2011

Investigating The Optimal Presentation Of Feedback In Simulation-based Training An Application Of The Cognitive Theory Of Multimedia Learning

Wendi Van Buskirk
University of Central Florida

INVESTIGATING THE OPTIMAL PRESENTATION OF FEEDBACK IN SIMULATION-BASED TRAINING: AN APPLICATION OF THE COGNITIVE THEORY OF MULTIMEDIA LEARNING

by

WENDI LYNN VAN BUSKIRK
B.S. University of Central Florida, 1998

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2011

Major Professor: Clint A. Bowers

ABSTRACT

There are many different training interventions that can be used in simulation based training systems (e.g., cueing, hinting, highlighting, deliberate practice, etc.). However, the most widely used training intervention in the military is feedback, most often presented in the form of a debrief. With advances in technology, it is possible to measure and diagnose performance in real-time. Thus it is possible to provide immediate feedback during scenarios. However, training systems designers should not consider the timing of feedback in isolation. There are other parameters of feedback that must also be considered which may have an impact on performance. Specifically, feedback content and modality may also have an impact on the appropriate timing of feedback and its' effectiveness in simulation training environments. Moreno and Mayer (2000) propose a cognitive theory of multimedia learning which describes how instruction is perceived and processed by a trainee. Using this theoretical framework, I investigate the optimal use of feedback while considering the interaction of feedback timing, content, and modality in scenario-based training environments.

In order to investigate the relationship between the timing, modality, and content of feedback, a 2 (immediate, delayed) X 2 (visual, auditory) X 2 (process, outcome) between-subjects design was used (a no feedback control condition was also included). Ninety participants were randomly assigned to the nine experimental groups. These participants performed a visual-spatial military task called the Forward Observer PC-based Simulation.

Results indicated that receiving feedback was beneficial to improve performance as compared to receiving no feedback. As hypothesized, during a visual-spatial task, auditory feedback presented during a scenario led to higher performance than visual feedback. Finally, while I did not support my hypothesis that an interaction between all three components of feedback would affect performance, it is promising that the pattern of results mirrored the hypothesized pattern. Theoretical and practical implications, as well as limitations of the current study and directions for future research are discussed.

ACKNOWLEDGMENTS

I would like to thank the numerous people who have provided not only professional but emotional support during the dissertation process. First and foremost, I would like to thank my major advisor, Dr. Clint Bowers and, my mentor, Dr. Gwendolyn Campbell. I truly would not have been able to reach this point without their support and guidance. Dr. Bowers has been supportive of me and my work even when others had given up on me. I truly appreciate that support and it has been a motivational source for me to get this thing done! Dr. Gwen Campbell and I have worked together for so many years. I highly regard her insights, guidance, and friendship. Wandering into NAWCTSD in 1997 and meeting her truly changed my life. I can't thank Dr. Campbell enough for all that she has taught me. I also extend my deepest appreciation to Dr. Salas and Dr. Mouloua who provided valuable feedback and forced me to think critically. Additionally, I would like to thank Alyssa Mercado for her tenacity in tracking down and practically forcing participants to show up for the experiment. This dissertation would not have been completed without her unwavering persistence.

I would also like to thank CAPT Dylan Schmorrow for funding this research under the Virtual Technologies and Environments (VIRTE) program. Additionally, I would like to thank Dr. Jim Sheehy who provided funding for participant compensation under the NAVAIR Independent Applied Research program. I also appreciate the efforts of John Kropewnicki and Alisa Marshall of Lockheed Martin who made changes to FOPCSIM in order to support the experimental manipulations. Thanks also go to Anthony Aakre for his assistance with data collection. I would also like to recognize Drs. Ami Bolton, Katherine Wilson, and Raegan Hoeft who I spent uncountable hours with discussing dissertation topics.

Lastly, I would like to thank my family and my husband, Chris. Thanks to Chris for being supportive and always providing encouragement during this process. Chris never complained any time when I kicked him out of the house so that I could get work done. I love you for that! Finally, my dad was my biggest supporter and was the driving force behind my education. I truly miss those weekly phone calls when he would ask how dissertation was going. He was the greatest inspiration to me and this dissertation is dedicated to his memory.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Problem Statement

Simulation is at the forefront of military training practices. Hundreds of different simulators can be found on military bases across the country ranging from full-scale flight simulators to pc-based procedural trainers. There are many cited advantages of simulation. Simulation allows for reduced cost by decreasing the amount of resources needed for training (Oser, Cannon-Bowers, Salas, & Dwyer, 1999). Additionally, simulation based training allows trainees the capability to practice situations that would be too dangerous (or costly) to perform in live environments (Rose et al, 2000). Finally, simulation provides trainers the capability to present scenarios that allow trainees to prepare for events that do not occur frequently (Corbett, Koedinger, & Anderson, 1997).

As simulation and virtual environment technology is becoming integrated in training practices, there has been a push to determine its actual training benefits. In the past, it was assumed that the simulators and virtual environments (VEs) were effectively and efficiently training the warfighter. However, this conclusion has been subject to debate (Rose et al., 2000; Salas, Bowers, Rhodenizer, 1998; Oser et al., 1999; Farmer, van Rooij, Riemersma, Jorna, Moraal, 1999). Part of the problem is that simulation development has taken a technology-centered approach versus a learner-centered approach (Mayer, 1999; Farmer, van Rooij, Riemersma, Jorna, Moraal, 1999). For example, when trying to improve simulators the focus usually is on physical realism (e.g., developing more realistic terrain, sea state, or weather models) instead of optimizing the training value.

Training systems designers should not solely focus on what technology can do, but how to improve performance and learning through the use of technology. By taking a learner-centered approach, a system designer must consider which training interventions should be used in conjunction with the simulation. There are many different training interventions that can be used in simulation based training systems (e.g., cueing, hinting, highlighting, deliberate practice, etc.). However, the most widely used training intervention in military simulation is feedback, most often presented in the form of a debrief.

It is generally believed that feedback is important for improving performance (Clariana, Wagner and Murphy, 2000; Kulhavy and Stock, 1989; Mory, 1992; Panasuk and LeBaron, 1999). However, the research support for this belief is not overwhelming. For example, in a meta-analysis, Kluger and DeNisi (1996) found that over 1/3 of feedback interventions actually weakened performance.

In fact, the military has focused on providing delayed feedback as a standard procedure after simulated training scenarios. However, with advances in technology, it is possible to measure and diagnose performance in real-time. Thus it is possible to provide immediate feedback during scenarios. The question remains: Is immediate feedback or delayed feedback presentation better for improving performance? However, training systems designers should not consider the timing of feedback in isolation. There are other parameters of feedback that must also be considered which may have an impact on performance. Specifically, feedback content and modality may have an impact on the appropriate timing of feedback and its' effectiveness in simulation training environments.

Moreno and Mayer (2000) propose a cognitive theory of multimedia learning which describes how instruction is perceived and processed by a trainee. Mayer (2001) suggests that instruction should be designed to allow the trainee to engage in active processing. In order to facilitate this, he suggests that temporal contiguity (i.e, timing), the ability to organize and integrate the information in sub-systems of working memory (i.e., modality), and the need to provide “process structures” (i.e., content) are important considerations. While this model has typically been used to present instruction in static, academic domains, I propose that the principles derived from this model can be applied to the development of instruction for dynamic, military training tasks.

In the sections that follow, I will present the cognitive theory of multimedia learning (Moreno & Mayer, 2000) as a theoretical framework for investigating the optimal use of feedback while considering the interaction of feedback timing, content, and modality in scenario-based training environments. Next, I will discuss the typical use of feedback in simulation based training, namely delayed feedback, and how that paradigm can be expanded. Finally, I will present an overview of the literature on the timing of feedback, content feedback, and modality of feedback.

Cognitive Theory of Multimedia Learning

Mayer (1999) has argued that when designing multi-media learning environments designers should take a learner centered approach versus a technology-centered approach. In other words, he argues that designers should focus on how to improve learning through the use of technology instead of focusing on what technologies can do. While his research has mostly centered on the use of animation in computer-based training, I believe his research and theories can be expanded and applied to other multi-media learning environments such as computer-based simulations.

Mayer defines multimedia as “the presentation of material using both words and pictures that is intended to enhance learning (p.2).” In other words, instructional designers have two main options for presenting instructional material to students: through the use of words and pictures. Mayer argues that their focus should be on using words and pictures in the right way to in order to enhance learning. Based on this premise, Mayer (2001) developed a cognitive theory of multimedia learning which is based on Baddeley’s theory of independent working memory sub-systems (Baddeley, 2000, 2001; Baddeley and Logie, 1999) and Wicken’s (1984) and Sweller’s (1988) theories on the limited capacities of these working memory subsystems. Specifically, his theory is based on the following assumptions: (1) that learner’s have independent auditory and visual working memory subsystems, (2) these working memory subsystems have a limited capacity, (3) learner’s have separate systems to process verbal and non-verbal information and, (4) “meaningful leaning occurs when a learner selects relevant information in each store, organizes the information in each store into a coherent representation, and makes connections between corresponding representations in each store. (p. 1)” Figure 1 presents Mayer’s cognitive theory of multimedia learning.

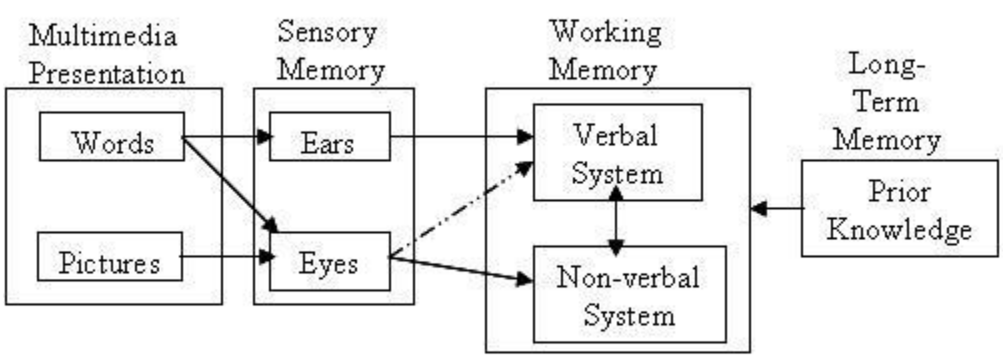


Figure 1: Cognitive Theory of Multimedia Learning (adapted from Moreno and Mayer, 2000)

The general idea is that instruction is presented via words and/or pictures and enters sensory memory through the learner's eyes and/or ears. In working memory (WM), words are processed in the verbal sub-system while pictures are processed in the non-verbal sub-system. Additionally, WM is used to temporarily hold and manipulate knowledge for active processing. This allows the leaner to construct knowledge in separate WM subsystems as well as integrate information from prior knowledge stored in long term memory. Mayer argues that the solid arrows from the ears to the verbal sub-system and the eyes to the non-verbal sub-systems represent the ideal cognitive processing for multimedia learning. This allows dual channel processing to occur which can increase the amount of information a student can process.

Based on research using this model, Mayer and Moreno (2000) have developed three principles of instructional design that foster active processing and are relevant for scenario based training environments. The first principle, Split-attention principle, states "Students learn better when the instructional material does not require them to split their attention between multiple sources of information (p.3)." More specifically, this principle suggests that students are more likely to split their attention when their sensory memory sub-systems are not taxed. If the same sensory channel is used to present information, the student may miss crucial parts of the instruction and, therefore, that information cannot be processed in WM.

The Modality principle states "Students learn better when verbal information is presented auditorily as speech rather the visually as on-screen text (p.4)." The rationale behind this principle is that the presentation of on screen text in addition to the animation or pictures being presented can cognitively overload students. Therefore, the use of auditory text can leave the visual, non-verbal channel free to process the pictures and animation.

Lastly, the Temporal Contiguity principle states "Students learn better when verbal and visual materials are temporally synchronized rather than separated in time (p. 5)." In other words, when visual and verbal information are presented at the same time, the learner is more likely to have both pieces of information in WM for active processing. If there is a temporal gap between the visual and verbal information, the learner is less likely to be able to make connections between the information in WM (Mayer, 2001). Designing training based on these three principles should allow the learner to actively engage in processing thus increasing the chances they will pay attention to the relevant information and be able to organize and integrate the information in WM.

Simulation as a Training Device

Simulation is at the forefront of military training practices. Thousands of different simulators can be found on military bases across the country ranging from full-scale flight simulators to pc-based procedural trainers. There are many cited advantages of simulation. Simulation allows for reduced cost by decreasing the amount of resources needed for training (Oser et al., 1999). Additionally, simulation replicates the operational environment and provides trainees the capability to practice situations that would be too dangerous (or costly) to perform in a live environment (Rose et al., 2000). Finally, simulation allows a trainer to present scenarios that allow trainees to prepare for events that do not occur frequently (Corbett, Koedinger, & Anderson, 1997). For example, researchers found that practice using an avionics troubleshooting simulation for 20-25 hours had an equivalent impact of 4 years job experience because participants were able to practice troubleshooting failures that might not happen regularly (Lesgold, Eggan, Katz, & Rao, 1992).

Despite these advantages, advances in simulation technology do not ensure that learning will take place. As noted by Farmer et al (1999), improving simulators as training devices has typically meant improving the physical representations of the environment. However, they suggest that training designers must also focus on instruction in order to optimize training effectiveness:

> Given the same training simulator, training results may differ widely depending on the way in which the training program has been designed and delivered. In this respect, the way in which instructional support is implemented is also an important determinant of training effectiveness and efficiency (p.63; Farmer, van Rooij, Riemersma, Jorna, Moraal, 1999).

Therefore, it is necessary to integrate appropriate training strategies, methods, and tools within these environments to achieve effective learning (Oser et al., 1997).

Scenarios that are presented to trainees are at the core of any training simulation. Therefore, the development of scenarios is a critical component to training. Oser and colleagues (1999) propose a framework to enhance the learning effectiveness of scenario-based training (SBT). In the SBT approach, the scenario is the curriculum and opportunities to practice different skills are presented during the scenario in simulations that mimic the operational environment. As part of the SBT process, trainees receive feedback, during a debrief, on the practice opportunities that were incorporated into the training scenario.

Training systems designers and developers have utilized the SBT process for several reasons (Oser, Cannon-Bowers, Salas, & Dwyer, 1999). First, SBT allows trainees to practice tasks that do not occur regularly in the operational environment (e.g., equipment failures). Additionally, SBT allows trainees to practice higher order skills such as problem solving and decision-making instead of focusing on procedural or declarative skills. Finally, SBT allows for systematic performance measurement in which a trainee's performance is objectively assessed against the events designed into the scenario. In fact, this approach has "resulted in improved performance in a variety of team training environments such as combat information centers, military air crews, and multi-service distributed teams (Oser et al., p.181)."

An important component of the SBT model is feedback. This model focuses on feedback delivered *after* scenario completion. However, the SBT process also easily lends itself to presenting feedback *during* a scenario. Therefore, it is possible to extend this model and propose that feedback may be more effective if it is delivered immediately, during a scenario.

Therefore, the focus of this dissertation is on how to optimally present feedback during the SBT cycle. However, there are many different aspects to consider when delivering feedback. This leaves one to wonder:

- When should feedback be delivered (during a scenario or after the scenario has been completed)?
- What content should be provided in feedback (velocity, normative, outcome, or process information)? Knowledge of the correct response (KCR)? Should trainees be required to answer until Correct (AUC)?
- How should the feedback be presented (Visual, auditory, or tactile modalities)?

With all of these different decisions, a framework to guide the optimal selection of feedback presentation is needed.

Mayer (1999) has argued that when designing multi-media learning environments such as simulation, designers should take a learner centered approach versus a technology-centered approach. In other words, he argues that designers should focus on how to improve learning through the use of technology instead of focusing on what technologies can do. Therefore, to derive theory-based, empirical guidance on how feedback should be delivered in scenario based training, I will rely on the Cognitive Theory of Multimedia Learning as a framework for my investigation.

Feedback

In general, feedback has been defined as “information about appropriateness of past performance (p.351, Ilgen, Fisher, Taylor, 1979)” or “any of the numerous procedures that are used to tell a learner if an instructional response is right or wrong (p. 211, Kulhavey, 1977).” As implied in the second definition and as can be seen in Table 1, there are many different forms of feedback that have been described in the literature. Despite the numerous types of feedback that have been investigated, all these types of feedback have the same underlying premise - presenting information to a trainee to help them learn correct behaviors.

Table 1: Definitions and Examples of Different Feedback Types

Feedback Type	Definition	Air Defense Warfare Example
Outcome	Provides knowledge of the results of one's actions (Kluger & DeNisi, 1996).	You were 80% correct when determining the intent of air contacts.
Normative	Provides an individual with information about his or her standing relative to others, but is not specific performance-related feedback (Smithers, Wohlers, & London, 1995).	You are at the 92nd percentile. 8% of operators were more accurate when identifying air contacts.
Velocity	The trainee's performance is compared only with his or her own prior performance on the task. The trainee can gauge the rate of progress at which a performance goal is being reached (adapted from Kozlowski et al., 2001).	You showed a 20% improvement on identifying targets from the last scenario.
Process	Conveys information about how one performs the task (not necessarily how well; Kluger & DeNisi, 1996).	You should take an air contact's speed and altitude into account when determining its intent.
Environmental	Provides information about the actual relationship between the cues in the environment and their outcomes (Balzer et al., 1994).	That target is hostile because it just shot a missile at ownship.

It is generally believed that feedback is essential to increase learning and performance (Ilgen, Fisher & Taylor, 1979; Locke and Latham, 1990). Further, Neth, Khemlani, & Gray (2008) argue that feedback provides a signal to monitor discrepancies between actual and desired states and initiate actions to correct mistakes during a scenario. Indeed, several meta-analyses have shown that presentation of feedback improves performance (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996). For instance, Kluger and DeNisi's (1996) meta-analysis results showed that providing feedback generally improved performance (ES = .41). However, these authors also found that in 33% of the studies in their review, feedback presentation actually decreased performance. In fact, Bangert-Drowns and colleagues (1991) reported similar findings in their meta-analysis. When reviewing the literature in this area, one finds that there are just as many studies finding positive effects of feedback and as there are studies finding negative and no effects of feedback. In fact, in her recent review of the literature Shute (2007) stated that despite 50 years of research, the feedback literature is riddled with conflicting findings and is one of the least understood features in instructional design. When trying to detangle the conflicting literature in this area, one must also consider feedback timing and feedback content.

Feedback Timing

Like the general feedback literature has a long history so does the literature on the timing of feedback. The main research question usually asks whether feedback should be given immediately after a student has responded to an item or some time (minutes, hours, weeks) after a task has been completed (Clariana, Ross, & Morrison, 1991)? Also similar to the general feedback literature, the results in the timing of feedback literature are also mixed and convoluted (Mason & Bruning, 2003). For instance, several researchers have found no significant differences between groups who received immediate or delayed feedback (Anderson, Kulhavy, & Andre, 1971; Clariana et al., 1991; Gaynor, 1981).

Generally, there are two schools of thought on the timing of feedback issue (Shute, 2007). Those who argue for the use of immediate feedback suggest that immediate feedback prevents errors from being encoded into memory (Bangert-Drowns et al., 1991). Those who argue for the use of delayed feedback suggest that providing immediate feedback on incorrect responses interferes with learning the correct way to do the task and errors made early on in learning are forgotten if delayed feedback is presented (Kulhavy & Anderson, 1972). In the following sections, I will review the theories and empirical support for both immediate and delayed feedback.

Theory and Support for Immediate Feedback Delivery

The theoretical perspective cited by proponents of immediate feedback is referred to as temporal contiguity. The notion is that when feedback is presented to a trainee in close temporal proximity of a response, the correct cue-strategy associations are strengthened and incorrect cue-strategy associations are weakened (Anderson, Corbett, Koedinger, & Pelletier, 1995; Corbett et al., 1997). This theory has its roots in Thorndike's (1913) law of effect:

> When a modifiable connection between a situation and a response is made and is accompanied or followed by a satisfying state of affairs, that connection's strength is increased. When made and accompanied or followed by an annoying state of affairs, its strength is decreased (p.4).

Likewise, Guthrie's (1935) contiguity theory suggests that all learning is a consequence of association between a particular stimulus and response. In fact, Bangert-Drowns and colleagues (1991) argue that consequences of behavior (i.e., a response) provide students information on "verification of retrieval accuracy, concept development, skill refinement, and metacognitive adaptation (p.214).” Additionally, they argue that feedback is most advantageous when it is used to immediately correct erroneous behaviors/responses. This has led to what is called the guidance hypothesis. The guidance hypothesis suggests that immediate feedback provides information about errors so that the learner may correct the errors on the next trial thus leading to improved performance (Schmidt, 1991). Indeed, results from meta-analyses lend support to both the temporal contiguity perspective and the guidance hypothesis.

Azevedo and Bernard (1995) conducted a meta-analysis to determine the effects of immediate feedback in computer-based instruction. Specifically, they analyzed effect sizes from 22 studies to compare the effectiveness of immediate feedback on immediate and delayed post-tests. They found that providing immediate feedback in computer-based instruction resulted in improved performance overall and showed an advantage of immediate versus delayed post-tests (effects of .80 and .35, respectively). Based on these results, they conclude “immediate delivery of a feedback message provides the best instructional advantage to the student (p.122).” While these results are promising, these authors compared immediate feedback delivery to no feedback control conditions. In other words, the relative effectiveness of immediate versus delayed feedback was not assessed in this meta-analysis.

Unlike the meta-analysis described above, Kulik and Kulik (1988) compared the effectiveness of immediate and delayed feedback on verbal learning tasks in the laboratory and in the classroom. Their review of 53 studies revealed that applied studies showed an advantage of immediate feedback over delayed feedback (average effect size = 0.28). However, results from laboratory studies showed the opposite effect (average effect size = -0.36) indicating an advantage of delayed feedback. There are issues to consider in light of the research proposed in this dissertation. First, the operational definition of immediate feedback is different than the one used in the proposed study – students received immediate feedback after the entire performance was completed, not after each individual response. Secondly, the tasks used in the meta-analysis encompassed declarative knowledge type tasks (e.g., list learning). Therefore, these results may not generalize to more complex military tasks that require decision-making and problem-solving. However, their results do show promise for the advantage of immediate feedback in more applied studies using more complex tasks.

Additionally, several studies have reported empirical evidence in support of the use of immediate over delayed feedback. Dihoff, Brosvic, and Epstein (2003) compared the effectiveness of delayed, immediate, and no feedback in a classroom setting. Specifically, students were presented with feedback on quizzes either 24-hours later (delayed), after each test item (immediate), or not at all (no feedback). The results showed that students who received immediate feedback preformed better on a 50 question final examination than both students who received delayed feedback or no feedback. Likewise, Guay, Salmoni, and McIlwain (1992) and del Rey and Shewokis (1993) found that participants who received feedback after every response performed better on acquisition trials than participants who received delayed feedback on a motor skills task. This phenomenon has also been shown using other types of tasks. For example, Corbett and Anderson (2001) found that presenting immediate feedback resulted in more efficient learning requiring less time on LISP programming lessons and Kirlik, Fisk, Walker, and Rothrock (1988) reported an advantage of immediate versus delayed feedback in a simulated military task.

In summary, there are numerous research studies that suggest immediate feedback presentation may be more beneficial than delayed feedback in training environments. Studies have shown that immediate feedback has improved performance effectiveness and efficiency across several different types of tasks - motor skills, verbal learning, programming, and command and control. In fact, Schooler and Anderson (1990) cite the following advantages of immediate feedback: “First, it increases the probability that relevant information will be in working memory. Second, it decreases the time spent floundering, focusing the subject’s attention on relevant information and decreasing time on task (p. 708)”. Additionally, proponents of the use of immediate feedback in instruction argue that it helps trainees learn appropriate cue-strategy associations. Indeed, this notion easily lends itself to military training problems stemming from the nature of their dynamic environments. Due to the rapidly changing environments, it is easy to see why feedback should be provided immediately after a response. It should be presented while the cue parameters in the environment remain unchanged and are still available to provide context for the feedback (Corbett, Koedinger, & Anderson, 1997). However, despite this compelling evidence, there are conflicting results and additional evidence that supports the delivery of delayed feedback over immediate feedback.

Theory and Support for Delayed Feedback Delivery

While there are no meta-analyses reporting that delayed feedback is better than immediate feedback, there are numerous studies that have reported this effect (Brackbill, Bravos, & Starr, 1962; Kulhavy & Anderson, 1972; Schooler & Anderson, 1990; Schmidt, 1991; Schmidt & Wulf, 1997; Sturges, 1969, 1972, 1978; Webb, Stock, & McCarthy, 1994) and an equal number of hypotheses have been presented to explain why delayed feedback presentation produces better performance than immediate feedback on post-test and retention scores (termed the delay-retention effect (DRE)). This first explanation for the DRE is coined the perseveration-interference hypothesis (Kulhavy and Anderson, 1972). This hypothesis suggests that when feedback is presented after an error, the incorrect response interferes with learning the correct answer. However, if a delay is presented, learners tend to forget their incorrect responses and there is a greater chance that a student will learn the correct answers from feedback (due to less interference). Kulhavy (1977) suggests that the perseveration-interference hypothesis is evident by research that has shown the probability of repeating an error on a posttest is significantly lower when feedback is delayed. To test this hypothesis, Webb, Stock, and McCarthy (1994) performed an experiment where participants took a multiple choice general knowledge test and either received feedback at the end of the test (immediate) or 24 hours later (delayed). They found that on a post-test, delayed participants were more likely to continue making correct responses and were more likely to correct errors originally made on the pre-test.

A second explanation presented for the delayed retention effect is that immediate feedback serves as a crutch. Schmidt (1991) suggests that learners come to rely on the presentation of immediate feedback to guide behavior and when it is removed, performance suffers. Schooler and Anderson (1990) further suggest that the dependence on the feedback "obscures the need to learn secondary skills necessary to perform the task without feedback" (p. 702). Schmidt and Wulf (1997) tested this hypothesis using a motor movement task. During pre-test, participants in their study either received feedback during performance of the motor task or received delayed feedback which compared their performance to the goal performance. Participants returned one day later for a post-test. Results showed that the immediate feedback group showed greater accuracy during pre-test performance. However, on the delayed post-test, the group receiving delayed feedback was more accurate and efficient in replicating the movement pattern than the group receiving immediate feedback during pre-test trials. Additionally, Schooler and Anderson (1990) investigated the effects of feedback timing in their intelligent tutoring system. While solving LISP programming problems, participants either received immediate or delayed feedback. Immediate feedback, which provided the correct answer, was given when the tutor detected an error during a step of the problem. Participants in the delayed condition received feedback at the end of the problem on the final solution. Their results showed that while participants went through the material 40% faster in the immediate condition, delayed participants made fewer errors on a post-test. In other words, while immediate feedback participants were more efficient, delayed feedback groups were more accurate.

The third explanation to support the delayed retention effect is that immediate feedback serves as an interruption by distracting attention from the task at hand (Schmidt & Wulf, 1997). Further, Schooler and Anderson (1990) and Schmidt (1991) suggest that the processing of immediate feedback competes for limited cognitive resources that are being used to perform the task and learning suffers as a result.

Conclusion

There are several reasons for the conflicting research results in the literature regarding the timing of feedback. First, the operational definitions of immediate and delayed timing are not used consistently in this body of literature. One researcher may define delayed feedback as 15 seconds after a response (Anderson et al., 1971) another study may define it as 24-hours (Sturges, 1978). Likewise, one researcher may define immediate feedback as feedback presented after each response (Sturges, 1972) while another research may define immediate feedback as feedback provided after the entire test or scenario has been completed (Webb et al., 1994). Therefore, it is possible that one researcher’s delayed feedback may be another researcher’s immediate feedback.

A second possible explanation for the conflicting results in the literature is that the appropriate timing of feedback may depend on the type of task being trained. For instance, there seems to be an advantage for the use of delayed feedback in programming and motor skills tasks, but an advantage for immediate feedback in declarative knowledge and decision-making tasks. Therefore, it remains an empirical question of whether or not these results will generalize to more complex military tasks.

Third, the content of the feedback used in the studies reported above has not been used consistently. Some studies have used feedback which provided information on the appropriateness of a student's response (e.g., the response was correct or incorrect; performance scores) while other studies have used feedback which provides information on how the student should perform the task. Therefore, it is hard to interpret patterns across these studies when the content of the feedback also differs.

Feedback Content

As previously shown in Table 1, there are many different forms of feedback that have been described in the literature. Most of the research in the feedback literature has focused on only two types of feedback: outcome and process. In the following sections, I will review the empirical support for both outcome and process feedback.

Delivery of Outcome Feedback

The type of feedback that is most often used in literature is outcome feedback. This type of feedback is defined as feedback which provides knowledge of the results of one's actions

(Kluger & DeNisi, 1996). Typically, outcome feedback takes one of the forms presented in Table 2 below.

Table 2: Types of Outcome Feedback

Feedback Type	Example
Knowledge of Response (KOR)	You were correct (or incorrect).
Knowledge of Correct Response (KCR)	Wrong, the correct answer is option "C"
Answer Until Correct (AUC)	Participants are not allowed to move on to the next item until they select the correct answer. By not being able to move on, they infer their answer is incorrect.
Percent Accuracy	You got 80% of the items correct.

Despite the widespread use of outcome feedback, there is not much empirical support for its effectiveness. For instance, Gaynor (1981) found no differences between outcome feedback and no feedback groups on a declarative knowledge task (i.e., matrix algebra). Additionally, within computer assisted instruction, several researchers found no advantage of providing outcome feedback over practice alone (Anderson, Kulhavy, & Andre, 1972; Roper, 1977; Morrison, Ross, Gopalakrishnan, & Casey, 1995). Similarly, while Kluger and DeNisi's (1996) meta-analysis results showed that providing outcome feedback generally improved performance, outcome feedback was shown to decrease performance in 1/3 of the studies they reviewed.

Despite these results, there has been some support for the use of outcome feedback over no feedback presentation. For example, Anderson, Kulhavy, and Andre (1971) performed two studies investigating the effectiveness of Knowledge of Correct Response (KCR) feedback during computer-based training on myocardial infarctions. Their results showed that groups who received KCR feedback outperformed no feedback groups. Additionally, Webb, Stock, and McCarthy (1994) found evidence for the use of outcome feedback when students were asked to learn a list of random facts. Despite the lack of evidence regarding its effectiveness and the existence of other types of feedback that may have a bigger impact on performance (i.e., process feedback), researchers and educators have continued to utilize this type of feedback.

Comparison of Outcome and Process Feedback

More recently, the use of process feedback over outcome feedback, especially for complex tasks, has been gaining favor. For instance, Earley, Northcraft, Lee and Lituchy (1990) state that “an individual who receives outcome feedback while performing an unstructured or complex task may make inappropriate adjustments (p. 89).” They further argue that trainee’s should receive feedback that focuses on the behavioral processes involved in performing a task rather than solely on the outcomes of behaviors. In other words, providing feedback on the processes and strategies of *how* to perform a task will have more of an impact on performance than feedback on performance outcomes. Likewise, Kluger and DeNisi (1996) argue that when performance is dependent on using overloaded cognitive resources, extra motivation provided by outcome feedback cannot “over compensate” to help the student perform better.

Bangert-Drowns et al. (1991) lend some initial support to the claims above. They reported in their meta-analysis that effect sizes were higher when process feedback was presented to students. However, this result was based only on 8 studies. Bisantz and Sharit (1993) compared the effectiveness of outcome and process feedback when using a natural language interface. On both the immediate and delayed post-tests, results showed that participants receiving process feedback were more efficient (completion score/good inputs) than participants receiving outcome feedback. Additionally, Gilman (1970) found that groups that received process feedback in addition to outcome feedback performed significantly better on a post-test than groups who received outcome feedback alone or no feedback on knowledge of general science concepts. Buff and Campbell (2002) compared the effectiveness of presenting outcome and process feedback in command and control military task. Their results showed that groups who received process feedback had significantly higher learning gains than groups who received outcome feedback or no feedback. Their results also showed that outcome feedback groups did not perform statistically better than no feedback groups. Lastly, Astwood, Van Buskirk, Cornejo, and Dalton (2007) compared the relative effectiveness of three different types of feedback (process, normative, and outcome) in a military decision-making task. Using planned comparisons, these authors found that participants who received process feedback outperformed participants who received normative, outcome, or no feedback on prioritization judgments. Similar to Buff and Campbell's findings, these authors also found that outcome feedback groups did not perform statistically better than no feedback groups.

Conclusion

While there is not an abundance of empirical support for the use of process feedback over outcome feedback, the results of the studies mentioned above show there is some promise for the use of process feedback. Indeed, in their review of the feedback literature, McLaughlin, Rogers, and Fisk (2006), argue that process feedback provides more instruction to trainees and feedback should relay what should have been done instead of simply told an error was made. Additionally, these authors suggest that "learning from feedback is a resource intensive activity. If researchers recognize that the use of feedback requires cognitive resources, it should be possible to predict how much and what kind of feedback is appropriate (p. 2626)." Likewise, Mayer (2001) also argues that instruction should be designed "in light of how the human mind works (p.4)." Therefore it may not just be the timing and content of the feedback that is important; instructional designers also need to consider the how the instruction will be cognitively processed by the trainee.

Feedback Modality

Information Processing

Generally, human information processing is thought to follow an input-process-output model. For example, in Stimulus-Central Processing-Response compatibility (S-C-R) schemes, the human information processing loop begins with sensory input (e.g., visual, auditory, haptic) or a *stimulus (S)*, which is then perceived and processed through working memory (e.g., verbal or spatial) or *central processing (C)*, and then *responded (R)* to by the human (e.g., vocally, manually), thereby completing the S-C-R processing loop (Wickens and Holland, 2000). It is believed that tasks demanding "verbal" working memory, such as interpretation of team communications, are thought to be best presented using auditory stimuli (i.e., speech), but could alternatively be presented via text. To optimize reaction time to such verbal information, a speech-based response is thought best. On the other hand, spatial information is thought to be best presented via graphics, but could alternatively be presented as sound localization or touch/motion. To optimize reaction time to such spatial information, a manual response is thought best.

Additionally, Multiple-Resource Theory (MRT; Wickens and Holland, 2000) suggests that individuals are more efficient in time-sharing tasks when different resources are utilized in terms of encoding perceptual stimuli (i.e., visual, auditory, haptic), processing codes (spatial, verbal), and responding (vocal, manual; see Figure 2). In other words, presenting spatial and verbal information through the visual and auditory channels respectively should result in an increased capability to multitask, as compared to presenting two visual tasks. I argue that immediate feedback presentation, during a simulation-based training scenario, can be thought of as second task that requires time-sharing. Therefore, the characteristics of the task as well as the characteristics of the feedback should be considered in order to optimize performance in simulation-based training scenarios.

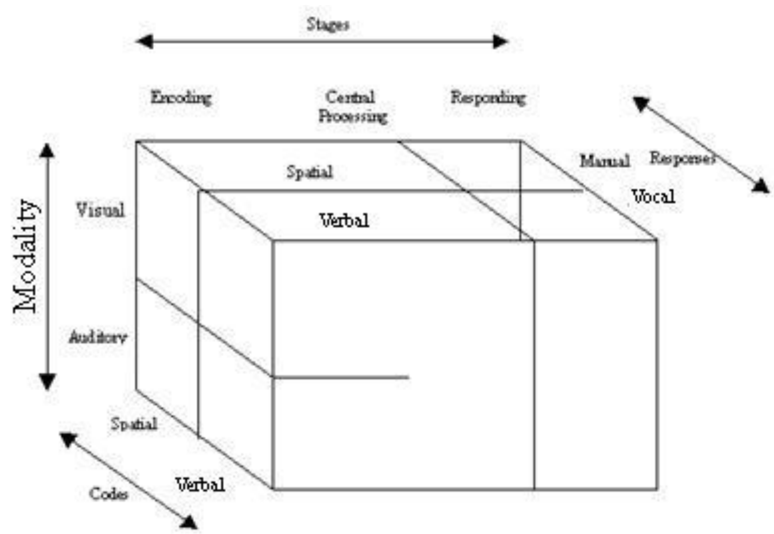


Figure 2: Multiple Resource Theory (Wickens and Hollands, 2000)

Research Support for Feedback Modality

As mentioned previously, Azevedo and Bernard (1995) performed a meta-analysis on 22 studies that investigated the use of feedback in computer based training. While they found that feedback improved performance over no feedback, they also found that modality of the feedback (e.g., graphics, verbal and auditory text) accounted for unique variance in the post-test data. Adepoju and Elliott (1997) also found support for the claim that modality of feedback mattered when presented during a second language learning task. In their study, participants were presented flashcards with different French words and they were to respond with the correct English word. Different types of feedback were presented after each response depending on the experimental condition participants were assigned to. The feedback conditions were (1) simultaneous written feedback (i.e., English and French words presented on the same flashcard), (2) written feedback (i.e., English word only presented on a flashcard), (3) pictorial feedback, and (4) aural feedback. Results showed that aural feedback presentation resulted in higher post-test performance than pictorial or both written feedback conditions. Additionally, they found that pictorial feedback resulted in higher post-test scores than both written feedback conditions. This suggests that visual feedback presentation may have interfered with the visual presentation of the stimulus (i.e., the flashcard). Therefore, the auditory feedback may have left available resources for encoding and processing the information. Likewise, Akamatsu, MacKenzie, and Habroucq (1995) investigated the use of sensory feedback using a target selection task. Their results showed that tactile and auditory feedback groups had reduced positioning times as compared to visual and no feedback groups on a visual, spatial task. Using a simulated driving task, Ferris, Hameed, Penfold and Rao (2007) found evidence for the use of haptic, spatial signals as attention aids when paired with visual verbal task. Further, they found a significant performance decrement when the haptic signal was paired with visual, spatial task. This finding provides evidence that it is not just the input modality, but also the processing code that you have to consider when designing instruction.

Despite the promising evidence above, Zolna and Catambrone (1997) found no evidence of this effect. Participants in their study received computer based training on the functioning of common objects (electric doorbell, refrigerator, etc.). They were presented with either verbal text or auditory narration in addition to animation and graphics on the subject matter. These authors found that replacing text with narration did not improve learning of the material. One potential explanation for this finding is that narration was presented via synthesized speech. Shneiderman (1988) has shown that listening and interpreting synthesized speech taxes working memory more than listening to human speech. Therefore, using synthesized narration might have caused an increase in cognitive workload instead of the decrease the authors were expecting.

The review of the literature above exemplifies the notion that presentation of feedback is not as simple as "should feedback be presented or not?" There are a complex mix of components that must be considered when designing feedback in scenario based training. For example, instructional designers need to consider the timing of feedback, the content of the feedback, as well as the modality of feedback. This requires instructional designers choose between several different methods and modalities to present feedback in the most effective manner. In order to guide instructional designers in developing optimal feedback, I propose that Mayer's (2001) cognitive theory of multimedia learning can be used as a framework to investigate and understand the complex parameters of feedback. Further, most simulated military tasks require visual-spatial processing (scanning tracks on a radar screen, flight simulators, tank location and identification). Therefore, in this dissertation, the experimental task will also be a visual, spatial task. Based on this, I posit that auditory verbal feedback would be most beneficial. More specific hypotheses will be described in the next section.

Hypotheses

Several meta-analyses, using over 679 effect sizes, have shown that presentation of feedback improves performance over no feedback presentation (Azevedo & Bernard, 1995; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996). Therefore, I hypothesize the following:

> *Hypothesis 1: Participants who receive feedback will outperform groups who do not receive feedback on a post-test.*

While there is not an abundance of literature on the use of process feedback, it stands to reason that providing strategies on *how* to perform a task better would be more beneficial to trainees than providing performance scores. Indeed, this logic may be particularly true when providing feedback on complex, dynamic, military tasks. Buff and Campbell (2002) and Astwood, et al. (2007), who compared the relative effectiveness of process and outcome feedback using different military decision-making task, found that participants who received process feedback outperformed participants who received no feedback. Further, authors from both studies found that outcome feedback groups did not perform statistically better than no feedback groups. Based on these findings, I hypothesize:

> *Hypothesis 2: Regardless of feedback modality, participants who receive delayed, process feedback will outperform groups who receive delayed outcome feedback on a post-test.*

Immediate feedback used during a scenario exercise can be considered a secondary task that requires time sharing. It requires the trainee to perceive and process the information presented in the feedback while performing the primary task. Therefore, the design of the feedback is an important consideration. Multiple-Resource Theory (Wickens and Holland, 2000) suggests that individuals are more efficient in time-sharing tasks when different resources are used. Further, these authors suggest that presenting spatial information through the visual channel and verbal information through the auditory channel should result in an increased capability to multitask.

Additionally, Wickens (2008) proposes a computational model which predicts the interference between time-shared tasks. Interference (I) is defined as “the sum of two components, a demand component (resource demand) and a multiple resource conflict component (degree to which overlapping resources are required)” (p. 451). The demand component for each task can be specified as being automated (D=0), easy (D=1), or difficult (D=2). While conflict (C) is defined as the extent to which the tasks share demands on common levels of the MRT Model (see Figure 2). Thus, interference can be calculated using the following equation:

$$I = (D_{task1} + D_{task2}) + C \qquad \text{Equation 1}$$

I applied this equation to the tasks required of participants in this dissertation who would receive immediate feedback. First, the demand for performing the experimental task ($D_{FOPCSIM}$) was assigned a “2” because the task is a complex, dynamic military task. Additionally, the demand for receiving and comprehending the feedback statements was also assigned a “2”. Finally, the only conflict that would occur in regards to the MRT model would occur on the modality level. Thus, the interference for auditory immediate feedback would be:

$$I = (D_{FOPCSIM} + D_{feedback}) + C = (2 + 2) + 0 = 4 \qquad \text{Equation 2}$$

While the interference for visual immediate feedback would be:

$$I = (D_{FOPCSIM} + D_{feedback}) + C = (2 + 2) + 1 = 5 \qquad \text{Equation 3}$$

Therefore, I hypothesize the following:

> *Hypothesis 3: When performing a visual-spatial task, participants who receive immediate, auditory feedback will outperform groups who receive immediate, visual feedback on a post-test.*

There are many variables to consider when designing and presenting feedback to trainees in military simulations. In this dissertation, I will be investigating the parameters of feedback timing, content, and modality. There are conflicting findings on which of the individual parameter is best for feedback presentation. However, I argue that these parameters should not be considered individually, but an interaction between the three parameters will result in the most optimal feedback.

One of the arguments against using immediate feedback is that it serves as a task interruption. For instance, Schooler and Anderson (1990) found that participants who received delayed feedback made fewer errors on a programming task than participants who received immediate feedback. These authors suggest that the processing of feedback competes for limited cognitive resources and that “if feedback were less disruptive, then they [participants] might return from the feedback episode with their goals intact (p. 707).” This suggests that feedback may not serve as a task interruption if it is designed to not be disruptive. Therefore, if you provide immediate feedback in a channel that is not overloaded, then immediate feedback may be more optimal than delayed feedback.

Additionally, proponents for the use of delayed feedback over immediate feedback argue that immediate feedback can serve as a crutch or that providing immediate feedback on incorrect responses interferes with learning the correct way to do the task (i.e., delayed retention effect). Again, I argue that both of these issues and findings are a result of the design of the feedback. For example, Guay et al. (1992) found that participants performed better during acquisition when presented with immediate feedback. However, when the feedback was removed (retention trials), participants in the delayed condition showed better performance than those in the immediate condition. Guay and colleagues suggested that the participants in the delayed feedback conditions had to generate their own solutions to problems. Thus they performed better on the post-test. Like many researchers in this area, these authors used outcome feedback. It makes sense that using active processing to determine how to correct mistakes while performing the task would lead to improved performance. However, I would argue that it may be possible to overcome immediate feedback serving as a crutch by providing process feedback instead of outcome feedback. Participants may not use process feedback as crutch because they will be provided with feedback explaining how to perform correctly on the next trial. Then, they can practice utilizing these processes during acquisition, which will lead to better performance during retention trials. Likewise, proponents of the delayed retention effect argue that delayed feedback is better than immediate because giving immediate feedback on incorrect responses interferes with learning the correct way to do the task. Further, when feedback is given after a delay, errors are forgotten and do not interfere (Kulhavy & Anderson, 1972). I would argue that this effect would not hold up especially if process feedback is given during complex tasks. For instance, if a trainee is given the process feedback immediately, they can maintain the context in working memory to help learn how to perform the task correctly on the next trial.

Further, Mayer (2001) argues that instruction should be designed to allow the trainee to engage in active processing. Like proponents of immediate feedback, he suggests that temporal contiguity is important. More specifically that the immediate presentation of feedback increases the chances that a learner will be able to hold corresponding visual and verbal representation of the same event in WM at the same time (or closer together in time). If there is a temporal gap between the visual and verbal information, the learner is less likely to be able to make connections between the information in WM. This argues for the use of immediate feedback. However, Mayer takes this one step further and suggests that training designers must also increase the chances that trainees pay attention to the relevant information and be able to organize and integrate the information in WM. When material is poorly designed, "learners must engage in irrelevant or inefficient cognitive processing" (p. 50; Mayer, 2001). If the same sensory channel is used to present information, the student may miss crucial parts of the instruction and, thus, cannot process that information in WM. Therefore, the modality of the feedback must be considered. Finally, Mayer also suggests that instructional designers must provide "process structures" which are cause-and-effect chains and consist of explanations of how some systems work. Therefore, the content of the feedback must also be considered. In summary, the best feedback message design is one that is presented during the task, in a processing channel that is not overloaded, and tells how the task should be done. Therefore, I posit the following:

Hypothesis 4(a): When performing a visual-spatial task, participants who receive immediate, auditory, process feedback will outperform groups receiving other combinations of feedback on a post-test

While I hypothesize that immediate, auditory, process feedback groups will perform best overall, I believe the relationship between the independent variables is more complex than the simplified hypothesis above. Therefore, I also hypothesize the following 3-way interaction below (see Figure 3):

> *Hypothesis 4(b): When considering performance on a post-test, if feedback is presented auditorily (i.e., in the underutilized channel), regardless of when it is presented, groups receiving process feedback outperform groups receiving outcome feedback. If feedback is presented visually (i.e., in the over utilized channel), the relative effectiveness of process and outcome feedback will depend on when it is presented. For instance, groups receiving outcome feedback will outperform groups receiving process feedback if it presented immediately. However, groups receiving process feedback will outperform groups receiving outcome feedback if it the presentation is delayed.*

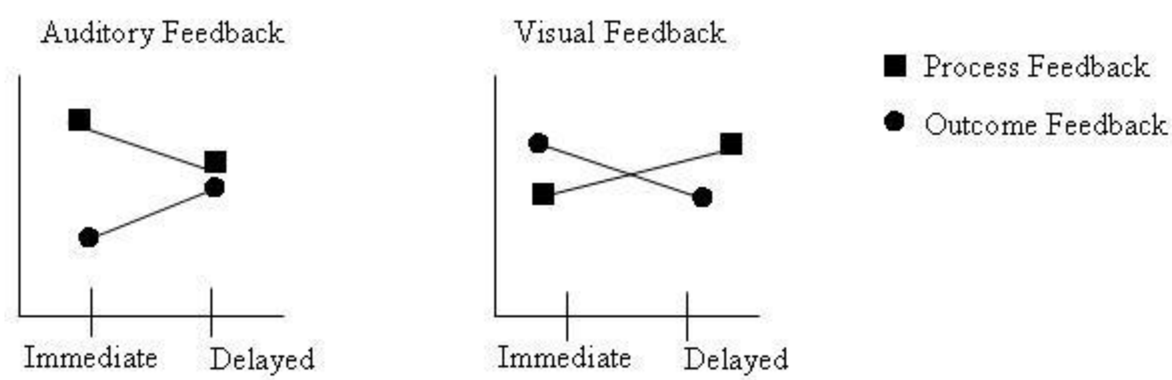


Figure 1: Hypothesized Three-way Interaction between Feedback Modality, Timing, and Content

Finally, more robust evaluations of military training effectiveness should be performed. The most optimal way to do this is to determine if trainees are still performing well after time has passed since their training. Since this is typically not feasible due to practical considerations such as time and budget constraints, Schmidt and Bjork (1992) suggest an alternative approach.

They argue that it is also important to determine if the skills acquired during training can generalize to novel situations or different contexts that were not present during training. For instance, research has shown that some types of feedback led to a decrease in performance during acquisition of a task. However, when presented with a transfer task, feedback presentation has shown to improve performance (Bisantz & Sharit, 1993; Schmidt, 1991). Since trainees receiving process feedback are provided with optimal strategies during acquisition, they will be more likely to apply these strategies under different contexts. However, this will only be true if the feedback is presented in a situation where there is free processing channel and in which the trainee can map the context with the right process. Additionally, it may be possible for trainees who receive outcome feedback during a scenario to engage in active processing on their own. Therefore, it may be possible to show the true power of different types of feedback by determining if they can increase performance under different conditions. Thus, I hypothesize the following:

> *Hypothesis 5: When considering performance on a transfer task, if feedback is presented auditorily (i.e., in the underutilized channel), regardless of when it is presented, groups receiving process feedback outperform groups receiving outcome feedback. If feedback is presented visually (i.e., in the over utilized channel), the relative effectiveness of process and outcome feedback will depend on when it is presented. For instance, groups receiving outcome feedback will outperform groups receiving process feedback if it presented immediately. However, groups receiving process feedback will outperform groups receiving outcome feedback if it the presentation is delayed.*

CHAPTER TWO: METHOD

Participants

Ninety participants (45 males, 45 females, M_{age}= 23.3 years, age range: 18-32 years) participated in the experiment. Participants were matched on gender to ensure equal numbers of males and females in each experimental group. They were recruited from Craigslist and received payment of $25 for their participation. Participants had no prior experience on the task and all participants were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" set forth by the American Psychological Association (2002).

Materials

Experimental Task

The testbed used in this dissertation was a modified version of the Forward Observer PC Simulator (FOPCSim). FOPCSIM was developed at the Naval Postgraduate School (NPS) and is a PC-based system that replicates the Call for Fire (CFF) task. A Call for Fire requires Forward Observers (FOs) locate targets and provide targeting information to a remote artillery unit. More specifically, the participant is responsible for determining the highest priority target based on a set of prioritization rules. Once the participant has determined the highest priority target, they identify the target, select the appropriate munitions, and enter the target's azimuth and distance information in the CFF template (see Figure 4). The participant enters this information using a standard mouse with a scroll wheel and keyboard. Prioritization rules, munitions tables, and descriptions of target types were provided to the participant and can be viewed in Appendix A.

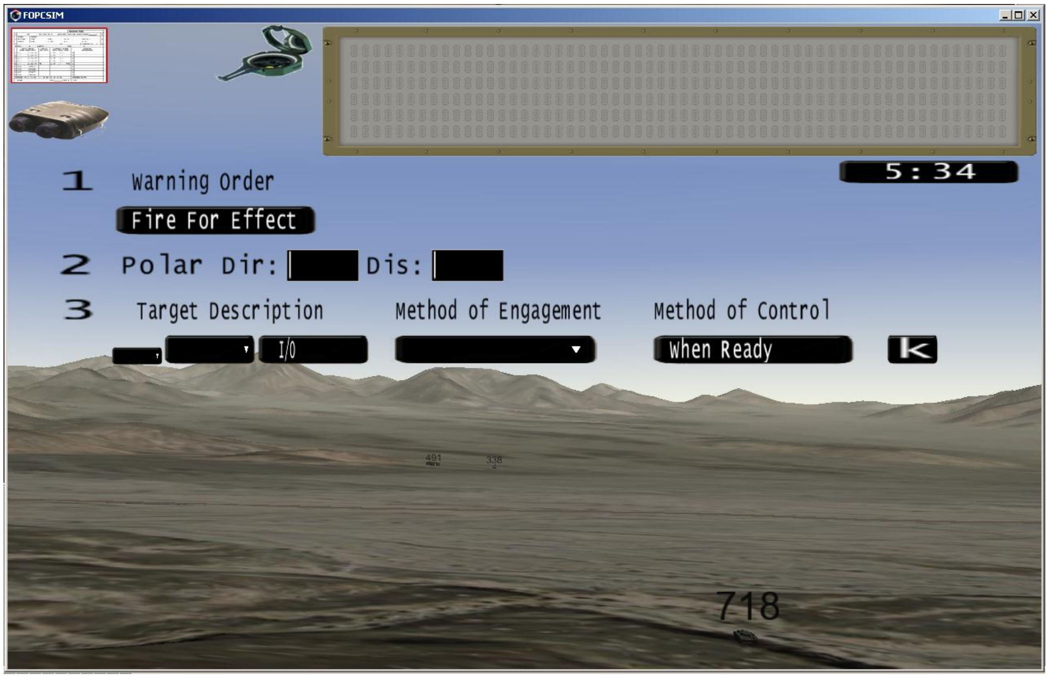


Figure 2: Screenshot of CFF Sheet

More specifically, during a scenario, participants determined which targets are of highest priority by scanning the environment in search of targets and by referring to prioritization rules provided. For example, a stationary target at 800 meters that is firing at the FO would be higher priority than a target moving toward the FO 600 meters away. Once the participant decides which target is the highest priority, he then uses the lensatic compass to determine the target's polar direction (see Figure 5). In order to determine the target identification and range, the participant uses the binoculars with laser range finder (see Figure 6). After the participant makes all of the assessments described above, they must enter the required information – target direction, distance, target number, target identification, and a munition selection – into the CFF sheet and then click on the "k" icon to transmit the target information to the artillery unit. After the CFF is complete and the rounds land, the trainee clicks on the “Continue” icon to clear the CFF sheet to begin a new mission.

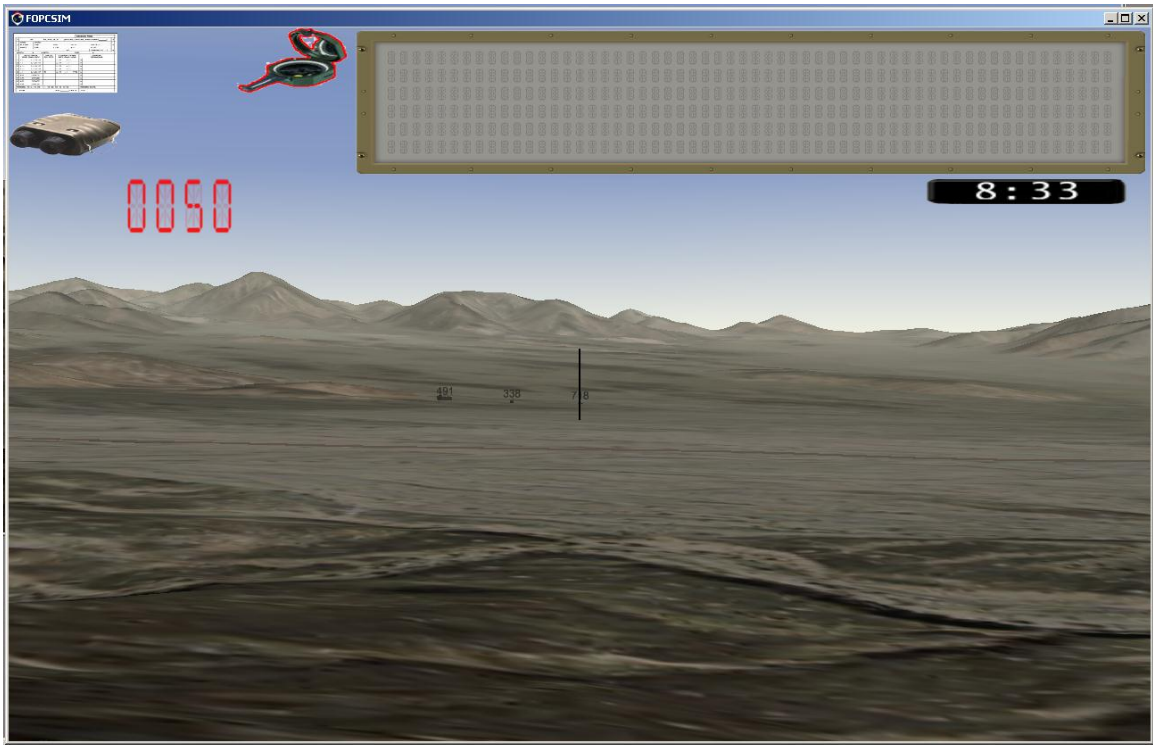


Figure 3: Screenshot of Obtaining Target Azimuth/Polar Direction

Figure 4: Screenshot of Binoculars with Laser Range Finder

The testbed also provides the capability to provide immediate (see Figure 7) and delayed feedback (see Figure 8) via text-based and audio-based messages.

Figure 5: Screenshot of Immediate Feedback Presentation

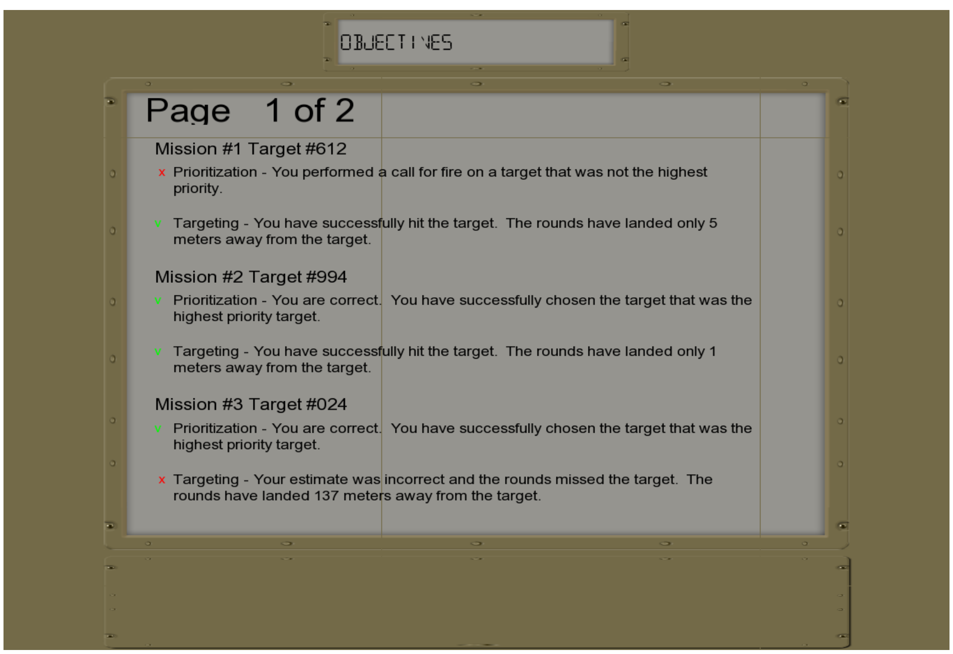


Figure 6: Screenshot of Delayed Feedback Presentation

Equipment

FOPCSim is a PC-based simulation and was run on an Alienware Area-51m 7700 laptop with an Intel® Pentium® 4 550J Desktop Processor, a NVIDIA GeForce™ Go 6800 video card with 256MB of DDR3 memory, and a 17” WideXGA screen. The simulation used a display resolution of 1280 x 800. A standard mouse with a scroll wheel was used for participants to interact with the simulation. Additionally, headphones were used for the participant to listen to environmental noise and auditory feedback.

Experimental Scenarios

A subject matter expert designed six scenarios for this experiment. The scenarios were developed to be representative of dynamic, complex tasks and care was taken to impose realistic demands on the operator throughout the scenarios. A second subject matter expert reviewed these scenarios and verified the cognitive load and realism of these scenarios. Further, the post-test scenario used in the experiment was rated by both SME’s as being the most realistic of the six scenarios. The training and testing scenarios were designed to be as similar as possible. For example, the scenarios contain 8 targets, one target engages the participant, and 3 targets move at 8 m/s. In addition to the training and testing scenarios, a transfer scenario was also developed. The transfer scenario was designed to present a novel situation to the participants. Specifically, the scenario contains 10 targets which move at 6 m/s and 10 m/s (instead of 8 m/s) and the placement of the targets requires participants to more actively scan the simulated environment to find the highest priority target.

Design Overview

In order to investigate the relationship between the timing, modality, and content of feedback, a 2 (immediate, delayed) X 2 (visual, auditory) X 2 (process, outcome) between-subjects design was used. Additionally, a no feedback control condition was also used. Participants were randomly assigned to one of the conditions presented in Table 3.

Table 3: Nine Experimental Conditions

Group	Timing	Modality	Content
A	Immediate	Visual	Outcome
B	Immediate	Visual	Process
C	Immediate	Auditory	Outcome
D	Immediate	Auditory	Process
E	Delayed	Visual	Outcome
F	Delayed	Visual	Process
G	Delayed	Auditory	Outcome
H	Delayed	Auditory	Process
I	No feedback		

Procedure

Upon arrival, participants were asked to read and sign the informed consent agreement and read the Privacy Act statement (Appendix J). Following completion of the informed consent form and demographic data form (Appendix C), participants completed questionnaires assessing individual differences that will be used in exploratory analyses (Appendix D). Then, individuals received training on FOPCSIM. The training contained information about the task they would be performing. Specifically, it reviewed information such as the rules of the game, the simulation screen, tools to use for obtaining data (e.g., direction and distance of targets), and symbology and buttonology. After the training, participants completed a knowledge quiz (see Appendix E) to assess if they paid attention during the training. If participants responded incorrectly to one or more questions, the experimenter reviewed those questions with the participant and discussed the correct answer. Next, the participant played a short demonstration scenario with experimenter coaching. (Note: The coaching was limited to helping find correct menus and pressing the correct buttons. The experimenter did not provide strategy information during this scenario.)

During the Experiment phase, participants were asked to complete three, 10-minute FOPCSIM scenarios in which they received either immediate, delayed, or no feedback. After each scenario, participants completed the Workload Manipulation Check (Appendix F) and Feedback Manipulation Check questionnaires (Appendix G). (Note: The no feedback group did not receive the Feedback Manipulation Check questionnaire.) When finished with the training scenarios, participants were given a 10 minute break. After the break, participants completed a testing scenario without feedback. Following the test scenario, participants played a transfer scenario also without feedback. Finally, participants were asked to complete a Feedback Reactions questionnaire (Appendix H), were debriefed on the purpose of the study (Appendix I), and were excused. The entire experiment took approximately 2.5 hours to complete. Table 4 presents the full experimental procedure and time estimates.

Table 4: Experimental Procedure

Activity	Time	Materials/Measures
Consent and Questionnaires	20 min.	Consent Form, Demographics Questionnaire, Individual Difference Questionnaires
Testbed Familiarization	40 min.	Testbed training, Demo Scenario, Knowledge Test
Break	10 min	N/A
Training Scenario	15 min	10 minute scenario, feedback and workload manipulation check questionnaires
Training Scenario	15 min.	10 minute scenario, feedback and workload manipulation check questionnaires
Training Scenario	15 min.	10 minute scenario, feedback and workload manipulation check questionnaires
Break	10 min.	N/A
Testing Scenario	10 min.	10 minute scenario
Transfer Scenario	15 min	10 minute scenario, Feedback Reactions Questionnaire
Debrief	10 min.	Debrief form
Total	150 mins.	

Experimental Manipulations

Timing of Feedback

For this experiment, the definitions of immediate and delayed feedback were operationalized such that immediate feedback provided feedback to the participant immediately following the completion of a CFF during the scenario or immediately following the missed opportunity for completing a CFF. Specifically, a missed opportunity for completing a mission occurs when a target reaches within a 100 meter radius around the FO, or at the completion of the ten minute scenario when targets remain that have not been neutralized. Delayed feedback was presented immediately following the conclusion of each training scenario and provided the exact same mission-by-mission CFF information that was delivered in the immediate feedback condition. These operational definitions of immediate and delayed feedback are consistent with military simulation-based training in which delayed feedback is typically presented in an After Action Review (AAR) moments after a training exercise has been completed.

Content of Feedback

Two types of feedback content were used in this experiment – outcome and process- as they are the most widely used in the feedback literature. Outcome feedback was operationalized as feedback that provides participant with the accuracy of their targeting and prioritization decisions. For example, “Incorrect. You did not disable the highest priority target.” Process feedback was operationalized as feedback that provides participants information on how to perform the task correctly. For instance, “Be sure to right click the mouse when using the laser range finder to determine a target’s distance.” Feedback templates can be found in Appendix J.

Modality of Feedback

Auditory or text-based feedback was presented to participants based on the condition to which they were assigned. Text-based feedback provided written information to trainee’s based on the content (i.e, either process or outcome) described above. Auditory feedback was presented to trainees using *.wav files of human voice recordings and is essentially a spoken transcript of the text-based feedback.

Measures

Performance Measures

Two aspects of performance served as the focus of training for the purposes of performance assessment and feedback: the degree of accuracy in identifying the highest priority target and the accuracy in determining the target's location. Participants were scored in each performance area at the conclusion of every mission and every missed opportunity for target neutralization. That is, when a moving target reaches within 100 meters of the participant's position, the participant has missed the opportunity to neutralize the target or perform any subsequent missions on that target as 100 meters signifies the no fire zone. While target identification and munitions selection are also sub-components of this task, several studies using FOPCSim have found ceiling effects on these sub-tasks (Astwood et al., 2008; Bolton, 2006). Therefore, they were not used for analyses in this dissertation.

Subjective Workload Measure

The Multiple Resource Questionnaire (MRQ) was used to measure subjective workload perceptions (see Appendix F). The MRQ is a 17-item questionnaire that measures workload within multiple cognitive resources (Boles, Bursk, Phillips, & Perdelwitz, 2007) during dual-task situations. For example, the respondent is required to rate the extent to which they used different processes, such as auditory linguistic processes or spatial attentive processes, during the task they performed. Responses on the MRQ were used as a manipulation check to assess that the appropriate working memory sub-systems are taxed in the experimental conditions. For example, I would expect to see that participants in the immediate auditory conditions rate a higher extent of usage on the auditory linguistic process than the visual phonetic process. While not all questions on the MRQ are relevant (e.g., facial figural process, tactile figural process), the scale was used as is. This will serve as another manipulation check to determine if participants were really paying attention to the questionnaire and/or CFF task.

Participant Reactions Questionnaire

Questionnaires designed to assess participant reactions to the training were adapted and slightly modified from Rhodenizer Van Duyne (2001) and Bolton (2006; see Appendix H). Participants completed this questionnaire at the end of the experimental session. The responses on this questionnaire were also used as a manipulation check. For instance, the questionnaire required respondents to rate items such as whether or not the feedback was easy to understand and if they "ignored and made no attempt to use the feedback."

Feedback Manipulation Check Questionnaire

A 4-item questionnaire was developed to determine if participants paid attention to the feedback they receive (see Appendix G). Participants assigned to feedback conditions, were required to answer questions such as "What information did the feedback provide when your munitions missed the target?" For a manipulation check, I expected participants to report information according to the condition to which they were assigned. For example, participants in the outcome conditions should report being told they received performance information (e.g., "I was told I hit/missed the target," "It told me I was doing well/poorly"). Likewise, participants in the process conditions should report being given information on how to perform the task better (e.g., "It told me to make sure I right click the mouse to get a target's range.")

Individual Difference Measures

A consistent finding in the literature is that males generally perform better on spatial ability tasks than females (Geary & DeSoto, 2001). The call for fire task used in this dissertation is a highly spatial task. Therefore, males may perform better on this task, overall, due to the male advantage in spatial abilities. Further, Bowers & LaBarba's (1988) research has indicated that right-hand motor activity interferes with spatial processing in females. Therefore, right-handed women may have a greater disadvantage when performing spatial tasks. Additionally, it is believed that some people are visual learners and some are verbal learners (Jonassen & Grabowski, 1993). In light of this literature, I included measures to use in exploratory analyses to determine if individual differences on gender, handedness, or the visualizer-verbalizer dimension affect performance.

There is some debate as to whether the visualizer-verbalizer dimension is a cognitive ability, cognitive style, or learning preference. Mayer and Massa (2003) performed a factor analysis using 14 different visualizer-verbalizer measures and found that each measure loaded on one of the factors mentioned above: cognitive style, cognitive ability, or learning preference. Based on their results, I used the questionnaire(s) that loaded most highly on each factor. To measure learning preference, I used the Multimedia Learning Preference Questionnaire (Mayer, 2002). To measure cognitive ability, I used reported SAT scores as well as the Verbal-Spatial Ability Rating questionnaire (Mayer & Massa, 2003). Finally, to measure cognitive style, I used the Verbalizer-Visualizer Questionnaire (Richardson, 1977) as well as the Santa Barbara Learning Style Questionnaire (Mayer & Massa, 2003). All visualizer-verbalizer measures can be found in Appendix D.

CHAPTER THREE: RESULTS

Manipulation Checks

Three manipulation checks were performed. First, analyses were performed to ensure that random assignment procedures worked and that all groups were equal at pre-test. Two manipulation checks were performed to verify that the experimental manipulations had their intended effect. Specifically, responses on the MRQ were analyzed to ensure that the experimental manipulations regarding working memory sub-systems had their intended effects. Lastly, analyses were performed to determine whether or not participants used and/or paid attention to the feedback they received.

Manipulation Check 1: Random Assignment

Analyses were performed to verify that groups did not differ on demographic variables such as age, GPA, video game experience (hours per week and type of game play), other game experience (e.g., word puzzles, picture puzzles, etc.), computer experience, or military experience. The means and standard deviations on the variables are presented in Table 5. A One-way Analyses of Variance (ANOVA) revealed that these groups were not significantly different: Age $F(8,88) = .82$, $p = .59$; GPA $F(8,81) = .57$, $p = .80$; hours per week playing video games $F(8,84) = .79$, $p = .62$; playing first-person perspective video games $F(8,87) = .21$, $p = .99$; playing third-person perspective video games $F(8,87) = 1.62$, $p = .132$; Solving word puzzles $F(8,87) = 1.09$, $p = .38$; Solving picture puzzles $F(8,87) = 1.48$, $p = .18$; computer experience $F(8,88) = .70$, $p = .69$; and military experience $F(8,88) = .86$, $p = .55$.

Table 5: Means and Standard Deviations on Demographic Variables

	A	B	C	D	E	F	G	H	I
Age	21.00	24.10	23.30	24.90	23.10	23.00	24.20	21.89	24.10
	(3.86)	(3.34)	(3.56)	(3.87)	(4.95)	(3.56)	(6.26)	(2.93)	(4.86)
GPA	3.32	3.02	3.22	3.25	3.16	3.28	3.34	3.26	3.20
	(.33)	(.52)	(.39)	(.28)	(.43)	(.37)	(.19)	(.29)	(.45)
Hours/week video games	1.80	5.45	6.44	6.20	2.78	2.67	1.11	5.44	4.05
	(1.70)	(9.17)	(10.85)	(9.83)	(1.97)	(3.12)	(1.36)	(8.80)	(4.98)
1st-person game experience	1.70	1.80	1.90	1.90	2.00	1.89	1.70	2.00	1.90
	(.48)	(.63)	(.74)	(.88)	(.94)	(.60)	(.67)	(.87)	(.88)
3rd-person game experience	2.10	2.60	2.70	2.50	2.70	2.00	2.40	2.44	2.30
	(.74)	(.52)	(.48)	(.71)	(.48)	(.50)	(.52)	(.73)	(.67)
Word puzzle experience	2.20	2.00	2.40	2.10	2.00	2.44	2.10	2.33	2.00
	(.42)	(.67)	(.52)	(.57)	(.67)	(.53)	(.32)	(.71)	(.00)
Picture puzzle experience	2.20	2.30	2.70	2.10	2.10	2.44	2.10	2.44	2.00
	(.63)	(.82)	(.48)	(.57)	(.57)	(.73)	(.32)	(.53)	(.47)
Computer experience	2.70	2.80	2.60	3.00	2.90	2.30	2.80	2.67	2.60
	(.67)	(.79)	(.70)	(1.49)	(.57)	(.48)	(.63)	(.50)	(.52)
Military experience	1.90	2.00	2.00	1.90	2.00	2.00	2.00	2.00	2.00
	(.32)	(.00)	(.00)	(.32)	(.00)	(.00)	(.00)	(.00)	(.00)

Note. 1st-person game experience, 3rd-person game experience, Word puzzle experience, and Picture puzzle experience: 1 = Not at all experienced, 2 = Somewhat experienced, 3 = Very experienced. Computer experience: 1 = No experience, 2 = Know a little (internet, Microsoft programs), 3 = Know quite a bit (e.g., other software, some programming), 4 = Expert (e.g., multiple software packages, multiple programming languages). Military experience is dummy coded where 1 = participant reported relevant experience and 2 = participant reported no experience.

To verify that random assignment produced groups that were equivalent on FOPCSIM performance, analyses on the demonstration scenario were performed. This scenario was selected because no feedback was presented to participants at this point in the experimental procedure and therefore could be used to determine that there were no differences in performance among the experimental groups. The means and standard deviations on prioritization and targeting performance are presented in Table 6. A One-way ANOVA revealed that these groups were not significantly different: Prioritization $F(8,87) = 1.22$, $p = .299$; Targeting $F(8,87) = .485$, $p = .864$.

Table 6: Means and Standard Deviations on Demonstration Scenario Performance

	A	B	C	D	E	F	G	H	I
Prioritization	59.40	28.70	62.60	56.70	54.22	43.89	41.00	41.60	45.70
	(26.66)	(35.76)	(19.15)	(25.69)	(20.10)	(37.09)	(35.15)	(40.03)	(31.54)
Targeting	47.60	37.90	41.60	51.00	58.56	36.67	54.00	40.80	45.00
	(29.11)	(37.38)	(20.06)	(39.00)	(27.79)	(37.30)	(37.18)	(33.66)	(31.76)

Note. Two participant's data (1 from Group E and 1 from Group F) were not recorded due to system logging errors and thus were excluded from this analysis.

<u>Manipulation Check 2: Experimental Manipulation</u>

Participant responses on the MRQ were used to assess whether the appropriate working memory sub-systems were taxed in the experimental conditions. Participants in the no feedback (Group I) and delayed conditions (Groups E-H) should report some usage of the following processes based on the nature of the FOPCSIM task: manual, short-term memory, spatial attentive, spatial concentrative, spatial emergent, spatial positional, and visual temporal. However, there should not be differences in participant's ratings on these processes. A One-way

ANOVA revealed that these groups were not significantly different on these processes: Manual Process $F(2,86) = .819$, $p = .444$; Short-term memory $F(2,86) = .290$, $p = .749$; Spatial attentive $F(2,86) = .253$, $p = .777$; Spatial concentrative $F(2,86) = .950$, $p = .391$; Spatial emergent $F(2,86) = .582$, $p = .561$; Spatial positional $F(2,86) = .378$, $p = .687$; Visual temporal $F(2,86) =$ .1.38, $p = .258$.

I expected to see participants in the immediate feedback conditions to have differences in their MRQ ratings due to the modality of feedback. For example, participants in the immediate auditory conditions (Groups C & D) should rate a higher extent of usage on the auditory linguistic process than those in the immediate visual conditions (Groups A & B). Likewise, participants in the immediate visual conditions should rate a higher extent of usage of the visual lexical process and short-term memory than those in the immediate auditory conditions. The means and standard deviations on auditory linguistic, visual lexical, and short-term memory usage ratings are presented in Table 7. Results showed that participants in the immediate auditory conditions did report higher usage of the auditory linguistic process, $t(38) = -1.81$, $p =$ .04. However, no differences between groups were found on usage of the visual lexical process ($t(38) = 0.26$, $p = .40$) . This result may have occurred because in both conditions (even the immediate auditory), participants needed to use visual lexical processes when filling out the CFF. Additionally, while the trends for mean ratings of short-term memory usage where in the expected direction, the differences between the immediate visual and the immediate auditory groups were not statistically significant, $t(38) = 131$, $p = .09$. It may have been that feedback content contributed to this result. Specifically, process feedback might have contributed to higher reports of short-term memory usage because it requires more cognitive resources to process the strategy information presented in that feedback compared to the simpler information presented in outcome feedback. Therefore, a t-test was performed to determine if immediate process groups (Groups B & D) reported higher short-term memory usage than immediate outcome groups (Groups A & C). Results showed that participants in the immediate process conditions [M=3.40 (.82)] did report statistically higher usage of short-term memory compared to participants in the immediate outcome conditions [M=2.89 (.62)], $t(38) = 1.71$, $p = .048$.

Table 7: Means and Standard Deviations on the Multiple Resource Questionnaire

	Immediate Visual (Groups A&B)	Immediate Auditory (Groups C&D)
Auditory linguistic	1.10 (1.20)	1.70 (.86)
Visual lexical	2.05 (1.27)	1.95 (1.14)
Short-term memory	3.30 (.80)	2.95 (.89)

The second manipulation check was used to assess whether or not participants used and/or paid attention to the feedback they received. Only 2 participants reported that they strongly agreed with the statement “I ignored and made no attempt to use the feedback I had received” on the Feedback Reactions Questionnaire (Appendix H). One participant was in the immediate, visual, process group and the other participant was in the immediate, auditory, process group.

After each scenario in which they received feedback, participants were asked to answer: “What information did the feedback provide when you selected a lower priority the target?” and “What information did the feedback provide when your munitions missed the target?” These free response questions were coded to reflect whether they matched or didn't match the content (e.g., process or outcome) of feedback they received. For example, when asked "What information did the feedback provide when you selected a lower priority the target?", one delayed, audio, process condition participant responded "Rounds completed." Additionally, on the feedback reactions questionnaire, participants were asked to respond to whether they received feedback during, after, or did not receive feedback after each scenario. Likewise, they were asked whether the feedback was presented with text, spoken or did not receive feedback. All feedback participants reported receiving feedback. Table 8 provides the frequency of mismatched reports of content, modality, and timing of feedback broken down by feedback condition. No participants misidentified the content, modality, and timing of feedback they received. Therefore, all participant data was used for analyses.

Table 8: Frequency of Incorrect Self-Reports on the Independent Variables

	Content	*Modality*	*Timing*
Immediate visual outcome	0	1	1
Immediate visual process	0	0	3
Immediate auditory outcome	0	0	2
Immediate auditory process	0	0	2
Delayed visual outcome	0	0	0
Delayed visual process	0	0	0
Delayed auditory outcome	2	2	1
Delayed auditory process	2	1	0

Hypothesis Testing

Means and standard deviations for post-test and transfer test prioritization and targeting performance can be found in Tables 9 and 10. To test Hypothesis 1, that participants who receive feedback outperformed groups who do not receive feedback on a post-test, two t-tests were performed on prioritization accuracy and targeting accuracy. To perform these analyses feedback was dummy coded where 1= no feedback and 2 = feedback. Regarding prioritization performance, results showed that participants who received feedback (M= 32.10, SD=25.67) outperformed those who did not (M=17.60, SD=16.30), $t(88) = -1.81$, $p = .03$. Additionally, results showed that participants who received feedback (M= 58.76, SD=24.55) outperformed those who did not (M= 41.10, SD=29.89) on targeting performance, t (88) = -2.09, p = .02. Therefore, Hypothesis 1 was supported for both dependent variables.

Table 9: Performance Means and Standard Deviations on Post-Test Prioritization and Targeting

	A	B	C	D	E	F	G	H	I
Prioritization	25.10	28.50	38.40	42.70	28.90	22.90	35.40	34.90	17.60
	(26.19)	(24.93)	(25.98)	(19.97)	(29.95)	(22.76)	(26.46)	(22.00)	(16.30)
Targeting	69.70	50.30	51.70	68.10	53.30	62.70	58.10	56.20	41.10
	(28.65)	(28.16)	(23.05)	(23.71)	(20.13)	(24.01)	(26.10)	(22.30)	(29.90)

Table 10: Performance Means and Standard Deviations on Transfer Test Prioritization and Targeting

	A	B	C	D	E	F	G	H	I
Prioritization	25.20	39.40	35.70	40.80	33.80	41.30	35.30	32.60	29.00
	(20.00)	(31.85)	(16.92)	(10.08)	(28.36)	(31.11)	(18.64)	(20.52)	(19.93)
Targeting	62.60	56.20	47.50	52.00	45.50	53.60	58.40	57.80	47.10
	(16.67)	(30.48)	(20.47)	(17.05)	(17.24)	(24.64)	(17.32)	(13.63)	(28.28)

Hypothesis 2 stated that regardless of feedback modality, participants who receive delayed process feedback will outperform groups who receive delayed outcome feedback on a post-test. Two t-tests were performed on prioritization accuracy and targeting accuracy. Regarding prioritization performance, results showed that participants who received delayed process feedback (M= 28.90, SD=22.64) were not significantly different that those who received delayed outcome feedback (M= 32.15, SD=27.71), t(38) = 0.41, p = .34. Additionally, results showed that participants who received delayed process feedback (M= 59.45, SD=22.80) were not significantly different than those who received delayed outcome feedback (M= 55.70, SD=22.82) on targeting performance, t (38) = -0.52, p = .30. Therefore, Hypothesis 2 was not supported.

Two additional t-tests were performed on prioritization accuracy and targeting accuracy to determine if receiving immediate auditory feedback improves performance more than immediate visual feedback (Hypothesis 3). Results showed that participants who received immediate auditory feedback (M= 40.55, SD=22.66) outperformed those who received immediate visual feedback (M= 26.80, SD=24.95) on post-test prioritization performance, t(38) = -1.83, p = .04. However, results showed that participants who received immediate auditory feedback (M= 59.90, SD=24.27) were not significantly different than those who received immediate visual feedback (M= 60.00, SD=29.39) on targeting performance, t (38) = 0.01, p = .49. Therefore, Hypothesis 3 was partially supported.

To test Hypothesis 4a, that participants who received immediate, auditory feedback will outperform all other feedback groups, two t-tests were performed on post-test prioritization and targeting accuracy. Regarding prioritization performance, results showed that participants who received immediate auditory feedback (M= 42.70, SD=19.97) out performed those who received other types of feedback (M=30.59, SD=25.03), t(78) = -1.73, p = .05. However, results showed that participants who received immediate auditory feedback (M= 68.10, SD=23.71) did not statistically outperform those who received other types of feedback (M= 57.43, SD=24.54) on targeting performance, t (78) = -1.29, p = .10. Therefore, Hypothesis 4a was only supported for post-test prioritization performance.

While Hypothesis 4a was only partially supported, exploratory analyses were performed to look for general effects of feedback type. More specifically, a series of t-tests were performed to determine which feedback groups were statistically different than the immediate auditory process feedback group (see Table 11) on post-test prioritization performance. These results showed that immediate auditory process groups performed significantly better than both immediate visual outcome and delayed visual process groups. Though, not statistically significant, the findings approach significance for the immediate visual process and delayed visual outcome groups. These results suggest a main effect of feedback modality on post-test prioritization performance.

Table 11: Comparison of Immediate Auditory Process Feedback Prioritization Performance to Other Feedback Groups

	t	*p*
Immediate visual outcome	-1.69	.05*
Immediate visual process	-1.41	.09
Immediate auditory outcome	-0.42	.34
Delayed visual outcome	-1.21	.12
Delayed visual process	-2.07	.02*
Delayed auditory outcome	-0.70	.25
Delayed auditory process	-0.83	.21

$df = 18$

Additionally, I hypothesized a 3-way interaction between content, timing, and modality of feedback (Hypothesis 4b). Specifically, I hypothesized that if feedback is presented auditorily (i.e., in the underutilized channel), regardless of when it is presented, groups receiving process feedback outperform groups receiving outcome feedback on a post-test. If feedback is presented visually (i.e., in the over utilized channel), the relative effectiveness of process and outcome feedback will depend on when it is presented. For instance, groups receiving outcome feedback will outperform groups receiving process feedback if it presented immediately. However, groups receiving process feedback will outperform groups receiving outcome feedback if it the presentation is delayed. To test this hypothesis, two hierarchical regressions were performed on each of the post-test performance variables – prioritization and targeting accuracy. In the first step, the individual predictors were forced into the equation. Then, interaction product term for the 3-way interaction was entered in the second block to determine if it improved prediction of performance beyond that of the individual predictors. Table 12 presents the hierarchical regression results including the R^2 and change in R^2 and Table 13 presents the unstandardized and standardized regression coefficients for each step in the model. As can be seen in Table 12, *R* was significantly different from zero at the end of Step 1. Table 13 shows that only modality of the feedback was a significant predictor of post-test prioritization performance (β = -.23, $t(76) = -2.11$, $p = .019$. The addition of the 3-way product term in Step 2 did not significantly improve R^2.

Table 12: Hierarchical Regression Results for Post-Test Prioritization Accuracy

Step	Variable(s) Added	F_{model}	df_{model}	p_{model}	R^2	ΔR^2	$F_{\Delta R^2}$	$df_{\Delta R^2}$	$p_{\Delta R^2}$
1	Modality, Content, Timing	3.73	3, 76	.007	.128				
2	Modality X Content X Timing	2.870	4, 75	.014	.133	.004	.375	1, 75	.271

Table 13: Regression Coefficients for Post-test Prioritization Performance

	Step 1			Step 2		
Predictor(s)	B	SE B	β	B	SE B	β
Modality	-5.75*	2.73	-.23*	-1.35	7.27	-.05
Content	-.15	2.73	-.01	10.13	15.96	.413
Timing	1.58	2.73	.06	-1.23	5.09	-.05
3-way Interaction				1.07	1.63	.48

*p<.05

The same procedure was used to test for the 3-way interaction using post-test targeting accuracy as the dependent variable. In the first step, the individual predictors were forced into the equation. Then, interaction product term for the 3-way interaction was entered in the second block to determine if it improved prediction of performance beyond that of the individual predictors. Table 14 presents the hierarchical regression results including the R^2 and change in R^2 and Table 15 presents the unstandardized and standardized regression coefficients for each step in the model. As can be seen in Table 14, *R* was not significantly different from zero at the end of Step 1 or Step 2. The addition of the 3-way product term in Step 2 did not significantly improve R^2. The 3-way interaction did not account for a significant amount of variance above that determined by the individual predictors for neither post-test prioritization performance nor post-test targeting performance. Therefore, Hypothesis 4b was not supported.

Table 14: Hierarchical Regression Results for Post-test Targeting Accuracy

Step	Variable(s) Added	F_{model}	df_{model}	p_{model}	R^2	ΔR^2	$F_{\Delta R^2}$	$df_{\Delta R^2}$	$p_{\Delta R^2}$
1	Modality, Content, Timing	.076	3, 76	.486	.003				
2	Modality X Content X Timing	.227	4, 75	.461	.012	.009	.679	1, 75	.206

Table 15: Regression Coefficients for Post-test Targeting Performance

	Step 1			Step 2		
Predictor(s)	B	SE B	β	B	SE B	β
Modality	.24	2.79	.01	5.91	7.43	.24
Content	-.56	2.79	-.02	12.68	16.31	.52
Timing	1.19	2.79	.05	-2.42	5.20	-.10
3-way Interaction				1.38	1.67	.62

Hypothesis 5 predicted that the same 3-way interaction between modality, timing, and content of feedback would improve prediction of transfer test performance over that of the individual predictors. To test this hypothesis, two additional hierarchical regressions were performed on each of the transfer performance variables – prioritization and targeting accuracy. In the first step, the individual predictors were forced into the equation. Then, interaction product term for the 3-way interaction was entered in the second block to determine if it improved prediction of performance beyond that of the individual predictors. Tables 16 and 18 present the hierarchical regression results including the R^2 and change in R^2 for transfer prioritization and transfer targeting, respectively. Additionally, Tables 17 and 19 present the unstandardized and standardized regression coefficients for each step in the model for transfer prioritization and transfer targeting, respectively. As can be seen in Tables 16 and 18, *R* was not significantly different from zero at the end of Step 1 or Step 2. The addition of the 3-way product term in Step 2 did not significantly improve R^2 for either transfer prioritization or transfer targeting. Therefore, Hypothesis 5 was not supported.

Table 16: Hierarchical Regression Results for Transfer-test Prioritization Accuracy

Step	Variable(s) Added	F_{model}	df_{model}	p_{model}	R^2	ΔR^2	$F_{\Delta R^2}$	$df_{\Delta R^2}$	$p_{\Delta R^2}$
1	Modality, Content, Timing	.473	3, 76	.351	.018				
2	Modality X Content X Timing	.359	4, 75	.419	.019	.000	.035	1, 75	.426

Table 17: Regression Coefficients for Transfer-test Prioritization Performance

	Step 1			Step 2		
Predictor(s)	B	SE B	β	B	SE B	β
Modality	1.18	5.17	.03	3.58	13.81	.08
Content	6.03	5.17	.13	11.63	30.29	.26
Timing	-.48	5.17	-.01	1.05	9.66	.02
3-way Interaction				-.29	1.55	-.14

Table 18: Hierarchical Regression Results for Transfer-test Targeting Accuracy

Step	Variable(s) Added	F_{model}	df_{model}	p_{model}	R^2	ΔR^2	$F_{\Delta R^2}$	$df_{\Delta R^2}$	$p_{\Delta R^2}$
1	Modality, Content, Timing	.045	3, 76	.493	.002				
2	Modality X Content X Timing	.103	4, 75	.490	.005	.003	.281	1, 75	.299

Table 19: Regression Coefficients for Transfer-test Targeting Performance

	Step 1			Step 2		
Predictor(s)	B	SE B	β	B	SE B	β
Modality	-.55	4.59	-.01	5.46	12.25	.14
Content	1.40	4.59	.04	15.43	26.87	.39
Timing	.75	4.59	.02	4.58	8.57	.11
3-way Interaction				-.73	1.38	-.40

Exploratory Analyses on Individual Differences

Exploratory analyses were also performed to determine if individual difference variables such as visualizer-verbalizer tendencies, handedness, and/or gender were correlated with performance on the FOPCSIM task. Correlations for the visualizer-verbalizer measures, gender, and DV's are presented in Table 20. With the exception of the Verbal-Spatial Ability Rating, none of the visualizer-verbalizer measures were correlated with prioritization or targeting performance. Further, the VSAR was only (positively) correlated with post-test targeting performance such that those who reported having a higher spatial ability than verbal ability performed better on post-test targeting (r= 0.33, p=.002). Additionally, gender was not correlated with any of the visualizer-verbalizer measures.

Due to the male advantage in spatial abilities that has been consistently reported in the literature, exploratory analyses were performed to determine if there were gender effects. Table 20 shows that gender was only significantly negatively correlated with post-test targeting (r= -0.40, p<.001) such that males tended to perform better on post-test targeting performance. *T*-tests were performed to compare performance of males and females on all 4 dependent variables. Results showed that males (M=32.64, SD=23.38) and females (M=28.33, SD=25.18) were not significantly different post-test prioritization, $t(88) = .842$, p=.201. Consistent with the correlational results, males (M=67.04, SD=22.70) performed significantly better than females (M=46.56, SD=24.46) on post-test targeting, $t(88) = 4.12$, $p < .001$. Males (M=37.93, SD=21.44) and females (M=31.64, SD=23.43) were not significantly different on transfer prioritization, $t(88) = 1.33$, p=.09. Finally, males (M=56.24, SD=15.88) and females (M=50.58, SD=25.18) were also not significantly different on transfer targeting performance, $t(88) = 1.28$, p=.11.

Lastly, *t*-tests were performed to compare performance of left-handed (n=9) and right-handed females (n=35) on all 4 dependent variables. Results showed that right-handed and left-handed females were not significantly different on either post-test or transfer prioritization performance. However, right-handed (M=51.91, SD=22.50) and left-handed (M=25.33, SD=22.46) females were significantly different on post-test targeting performance, $t(42) = -3.16$, $p=.008$. Right-handed (M=53.94, SD=24.05) and left-handed (M=33.89, SD=23.12) females were also significantly different on transfer targeting, $t(42) = -2.25$, $p=.04$. There were only 3 males who reported being left-handed, therefore, these analyses were not performed due to the low number of data points.

Table 20: Correlation Matrix for Visualizer-Verbalizer Measures

Measure	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. Gender	-										
2. SAT Verbal	-.03	-									
3. SAT Math	-.09	.16	-								
4. SBLQ	-.06	-.11	.10	-							
5. VVQ	.06	-.21	-.24	.04	-						
6. MMLPQ	-.05	.02	.05	.27**	.17	-					
7. VSAR	-.14	-.26	.13	-.15	.22*	.14	-				
8. Post-test Prioritization	-.09	.10	.08	.04	.00	.05	.15	-			
9. Post-test Targeting	-.40**	.08	.13	.13	-.05	-.04	.33**	.19	-		
10. Transfer Prioritization	-.14	-.01	.08	.08	-.03	.02	.15	.37**	.17	-	
11. Transfer Targeting	-.14	.09	-.10	.21	.08	-.05	.18	.21*	.37**	.32**	-

**p < .05, **p < .01*

CHAPTER FOUR: DISCUSSION

Overall, the results of this dissertation provide support for the use of feedback in dynamic military tasks. Consistent with meta-analyses by Azevedo and Bernard (1995), Bangert-Drowns, Kulik, Kulik, and Morgan (1991), and Kluger & DeNisi (1996), I found that feedback groups outperformed no feedback groups on both prioritization and targeting performance. Additionally, I intended to add to the literature in support of process feedback over outcome feedback. However, I did not find statistically significant differences between process and outcome groups on either dependent variable. My results also showed that modality of feedback is important to consider especially when feedback is presented during task performance. Consistent with Azevedo and Bernard (1995), I found partial support that immediate auditory feedback groups had higher prioritization accuracy than immediate visual groups. However, there were no differences between groups on targeting accuracy. Further, I found that immediate, auditory, process participants outperformed all other feedback groups on the prioritization task. Unfortunately, this finding was not replicated when using targeting accuracy as a dependent variable.

I also set out to show that different feedback parameters such as timing, content, and modality should not be considered in isolation and used the Cognitive Theory of Multimedia Learning as a framework to determine the most optimal feedback presentation. Based on this framework, I hypothesized that an interaction between the three parameters would result in the most optimal feedback presentation. Unfortunately, the 3-way interaction was not supported for post-test nor transfer performance.

Figures 9 and 10 present the pattern that was found in the post-test data. Regarding prioritization performance (Figure 9), when compared to the hypothesized pattern (see Figure 3), you can see that the visual delayed feedback results were opposite of the hypothesized direction. Overall, this graph shows the main effect of modality of feedback as well as the finding that immediate auditory process participants had the highest prioritization performance.

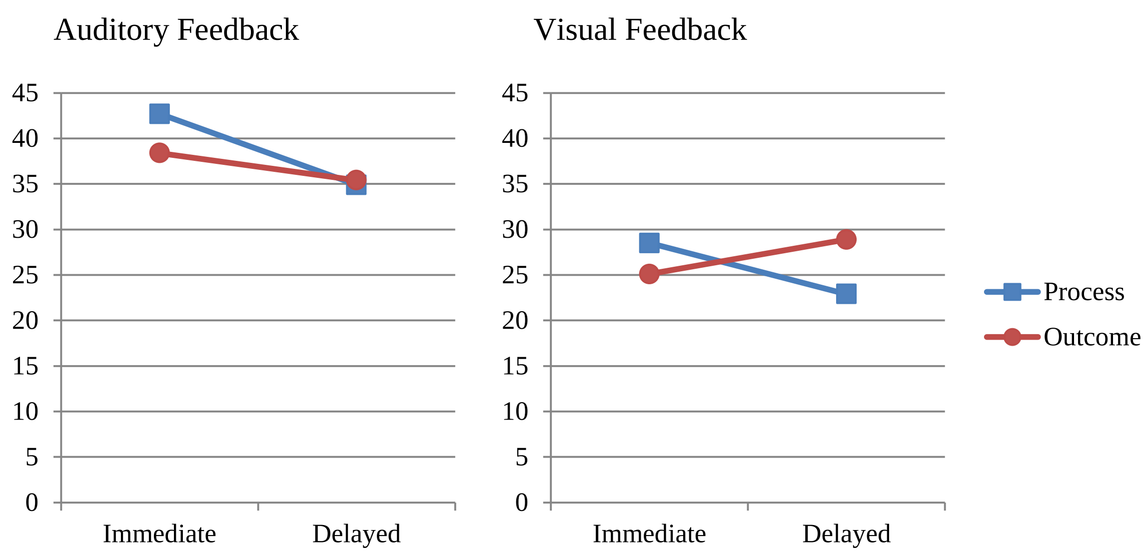


Figure 7: Three-Way Interaction on Post-Test Prioritization.

Regarding targeting performance (Figure 10), while not statistically significant, it is promising to see that the pattern of results is consistent with the hypothesized pattern that was presented in Figure 3. One exception was that I expected to see immediate auditory process participants would have had higher targeting accuracy than immediate visual outcome participants.

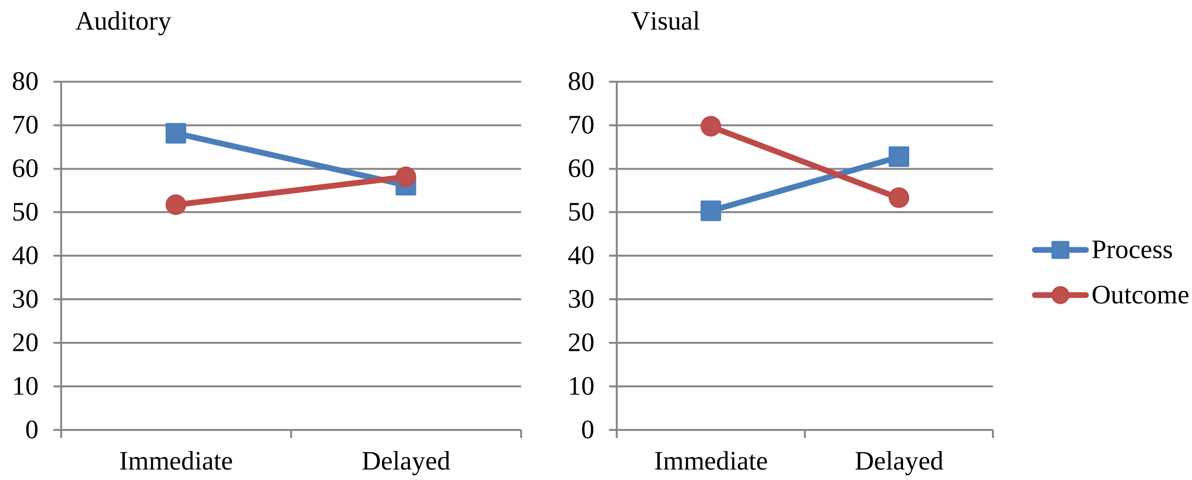


Figure 8: Three-Way Interaction on Post-Test Targeting

There are several explanations for the experimental findings. First, I will consider why process groups did not out perform outcome groups. I hypothesized that providing strategies on *how* to perform a task better would be more beneficial to trainees than providing knowledge of results. It may have been the case that the outcome feedback was, unintentionally, more helpful to participants than the process feedback for targeting performance. As is shown in Appendix J, when participants missed a target, the outcome feedback statements provided information on how far off the range was (e.g., "You were X meters from the target") while process feedback statements provided the following information: "Your munition landed behind/in front of the target. It will be easier to hit a moving target if you enter all other information in the CFF sheet BEFORE checking and entering the range." Providing the exact number of meters they were off on their range estimates may have helped participants more precisely determine where their errors were occurring versus simply telling them they over- or under-estimated and enter the range in the CFF sheet last. Lastly, while I developed the process feedback to target the most frequently made errors; it may be that process feedback not focusing on correct type of performance error on each individual CFF the participant made. For example, process feedback for incorrect targeting might have said “Your munition landed behind the target. It will be easier to hit a moving target if you enter all other information in the CFF sheet BEFORE checking and entering the range.” But, in reality, the participant may have just made a typo when entering range into the CFF sheet.

Additionally, several hypotheses were only supported for prioritization performance and not targeting performance. One explanation for this finding is that the prioritization task required more spatial processing than the targeting task. For instance, participants were required to visually scan and pick out objects in space. Additionally, they were required to use the shape of objects to determine the type of target (e.g., bunker or tank) in order to determine if it was a higher priority over the other targets. Alternatively, the targeting task required entering data about the targets in the CFF sheet and performing mathematical operations (on moving targets only). Therefore, it may have been more likely for modality of feedback to have a larger affect on prioritization performance due to the higher spatial processing required to perform that sub-task.

Another possible explanation for the lack of findings on targeting performance is that environmental feedback is provided on that sub-task. For example, on the targeting sub-task, participants can see where the munitions land and, if a target is hit, it appears to smoke. While this is a realistic feature, it may have been a potential confound especially for delayed feedback groups. Specifically, the environmental targeting feedback provided a double stimulus exposure (Kulik & Kulk, 1988). In other words, participants in the delayed feedback groups received feedback on the accuracy of their targeting calls twice - once during the scenario from the environmental feedback and a second time at the end of the scenario. On the other hand, participants had no way of knowing whether or not they correctly prioritized until they received the explicit feedback message.

Finally, as can be seen in Tables 10 and 11, there was a large amount of variance in the performance data. This large variability in the scores may have decreased the probability of finding significant results. To determine the sample size needed to have sufficient power to reject the null hypothesis, I performed two power analyses using the effect sizes from the hierarchical regression analyses on post-test prioritization and targeting. Using the procedures described by Cohen and Cohen (1983) power was set at .80 and alpha was set at .05. Using Cohen's f^2 for hierarchical regression, effect sizes were set at .006 for prioritization and .009 for targeting. The power analysis revealed 202 participants per condition for prioritization and 135 participants per condition for targeting would have been needed. The power analysis calculations are presented in Table 21.

Table 21: Power Analysis Calculations Using Effect Sizes from Current Study

Power Analysis Equation $n^* = L/f^2 + k + 1$	Prioritization	Targeting
$\boldsymbol{L}$	10.90	10.90
$\boldsymbol{f^2 = R^2_{ab} - R^2_{a} / 1 - R^2_{ab}}$	.006	.009
$\boldsymbol{K}$	3	3
$\boldsymbol{n^*}$	1821	1215

Considering performance on the transfer task, a 3-way interaction was not found for either transfer prioritization or transfer targeting. Further, none of the individual feedback components were significant predictors of transfer performance. In addition to the issues listed above, the transfer task was also more difficult and participants performed worse on the transfer task which may also explain the lack of findings.

I also performed exploratory analyses to determine if individual differences had an effect on FOPCSIM performance. Despite the finding that males generally have higher spatial ability than females (Geary & DeSoto, 2001; Halpern, 2000), I found that males only performed better than females on post-test targeting. This finding may have occurred not because of a spatial advantage but because of differences in mathematical ability from this sample. The only measure of math ability in this dissertation was self-reported Math SAT scores. A *t*-test was performed to see if males reported higher SAT math scores than females (M=602 vs M=584) and no statistical differences were found.

Generally, it is believed that people who score high on spatial ability, visual cognitive style, and visual learning preference will perform better on spatial tasks (Mayer & Massa, 2003). With the exception of the Visual-Spatial Ability Rating, my results didn't support this finding. This may be due to the fact that the FOPCSIM tasks used in this dissertation were not purely spatial; it contained a verbal component as well. Previous research that has found this result used purely spatial task stimuli such as paper folding tasks.

Limitations and Future Research

Several limitations to the current study should be noted. First, I failed to find that process feedback was better than outcome feedback. As mentioned previously, this could have been due to the way that the outcome feedback statements were written. Therefore, future research should consider assessing the relative effectiveness of a combined process and outcome condition, process feedback alone and outcome feedback alone conditions.

Additionally, it may have been that the error the participant made at that time was not addressed in the feedback (e.g., a making a typo). Furthermore, it was important to keep feedback statements relatively short (especially for immediate feedback groups). Therefore, while participants could have made multiple errors, the feedback statements only addressed one specific error on targeting and prioritization. Though I tried to compensate for this by having different feedback based on if the target was moving versus stationary, future research should address whether intelligent adaptive instruction would prove more useful to target errors on a more detailed, case by case, basis.

I also failed to find a significant 3-way interaction. As is typical of laboratory training experiments, the participants only received approximately 70 minutes of training. This amount of training is relatively short and it may be possible that effects of the variables would have been uncovered if additional scenarios training scenarios were added. As mentioned previously, there was a large amount of variance in the performance data which may have decreased the probability of finding significant results. Therefore, several changes to the experimental design could be made in the future in order to increase the probability of finding significant results. First, an increase in the sample size is needed. However, increasing a sample size large enough may not be practical (N=1200-1800). Additionally, it would be possible to decrease the variability in the scores by using a simpler experimental task. Alternatively, using participants that had characteristics closer to military populations (such as ROTC students) may have also been useful. Finally, more experimental control could have been exerted. Though realistic, in order to eliminate environmental feedback being provided on targeting performance, the experimental testbed could have been changed by not showing the rounds landing on the target.

Future research should also consider assessing the relative effectiveness of providing other types of feedback that were not used in this dissertation, namely velocity and normative feedback. Initial work on assessing the relative effectiveness of different types of feedback content showed an interaction between gender and feedback content (Landsberg, Van Buskirk, Astwood, 2010). Using a similar task, these authors found that process feedback was more beneficial to females while velocity feedback was more beneficial to males. However, the feedback provided in this study was summary-based and delivered post-scenario.

While assessment of the visualizer-verbalizer construct was not the aim of this dissertation, future research should also focus on more robust experiments to investigate the benefits of incorporating visualizer-verbalizer cognitive styles or learning preferences into training practices. Pashler, McDaniel, Rohrer, and Bjork (2009) suggest that participants must be divided into experimental groups based on their learning style scores then randomly assigned to experimental conditions. However, before that can be accomplished, better measures of visualizer-verbalizer are needed. For instance, the Visualizer-Verbalizer Questionnaire, while widely used, is known for its low reliability (Leutner & Plass, 1998). Additionally, the visualizer-verbalizer construct may not be a dichotomy. It may not be as simple as one person is a visualizer and another is a verbalizer. What if a person scores high on both? To complicate matters further for training practitioners, how would you design training for a trainee who has a visual cognitive style, but a verbal learning preference?

Finally, the Cognitive Theory of Multimedia Learning has been traditionally used to provide instruction in classroom settings in domains such as natural science (e.g., lightning formation) and mechanics (e.g., functions of brakes and pumps). I used this theoretical framework to determine the most optimal feedback presentation in a dynamic, multimodal, military task. The results of this dissertation found empirical support, on the prioritization sub-task, for the modality and split-attention principles which specifically address sensory memory component of the theory. For example, when feedback was presented during a scenario such that verbal information was presented via speech rather than text, it did not require participants to split their attention between the visual and auditory channels. Thus, participants in the immediate, auditory feedback groups had higher performance.

However, I failed to find support for the targeting sub-task which may have been due to over-generalization of the theory. For instance, Mayer (2001) states that words are processed in the verbal sub-system while pictures are processed in the non-verbal sub-system. Further, that when visual and verbal information are presented at the same time, the learner is more likely to have both pieces of information in WM for active processing. Since the targeting task contained a visual, non-verbal component (projecting the future location of the target), I hypothesized that immediate auditory, verbal feedback would result in higher performance. However, as mentioned previously, the targeting task also required interacting with the CFF sheet and using the targeting formula which requires cognitive processing in the verbal WM sub-component in addition to processing in the non-verbal WM sub-component. Therefore, future research should address expanding the theory to deal with instances when tasks require using both the verbal and non-verbal sub-components simultaneously.

CHAPTER FIVE: CONCLUSION

In summary, the current experiment investigated three parameters of feedback (timing, modality, and content) on performance. Results indicated that receiving feedback was beneficial to improving performance on a simulated, military task. Additionally, this dissertation highlights the importance of considering the modality of feedback. As hypothesized, during a visual, spatial task, auditory feedback presented during a scenario led to higher performance than visual feedback. Finally, while I did not support my hypothesis that an interaction between all three components of feedback would affect performance, it is promising that the pattern of results mirrored the hypothesized pattern. Therefore, I believe that future research investigating the three-way interaction is warranted.

The current study also has theoretical and practical implications. First, I contributed to the feedback literature by extending the Cognitive Theory of Multimedia Learning framework to a complex, multimodal, military task. Further, this theoretical framework proved to be useful for deriving theory-based, empirical guidance on how feedback should be delivered in scenario based training environments. Additionally, I extended the relatively scant literature on the modality of feedback. Lastly, the current research confirms the need for instructional designers to take a learner-centered approach and consider the different parameters of feedback, especially the modality of feedback, when designing training systems. The military makes extensive use of simulation based training and providing sound instructional support, based on empirically validated principles, within those systems would provide a better investment.

APPENDIX A: BRIEFING PACKET

Briefing Packet

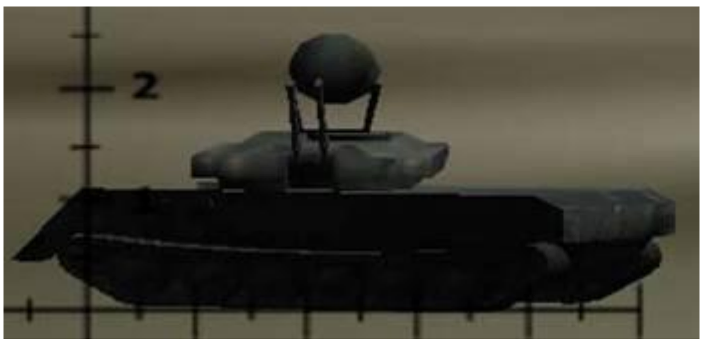
Target Prioritization Rules

Target missions must be conducted in accordance with the following rules:

1. Neutralize targets engaging your position.
2. Neutralize the nearest moving target within 100-2,000 meters from your position.
3. Neutralize the nearest stationary T-72.
4. Neutralize the nearest stationary ZSU.
5. Neutralize the nearest stationary bunker.
6. Do not neutralize targets beyond 2,000 meters from your position or within 100 meters of your position.

Target Types

Heavy armor vehicle – T-72 (tracked wheels and a long barrel gun on top)



Light armor vehicle – ZSU 23-4 (tracked wheels and a radar dish on top with lots of small barrels)


Ammo Bunker (square structure used for storing ammunition)

Munition Effectiveness on Target

Munition Effectiveness on Target

		Munition Types		
		ICM	VT	HE/Quick
Target	T-72	**100%**	10%	10%
Types	ZSU	10%	**100%**	10%
	Bunker	10%	10%	**100%**

APPENDIX B: DEMOGRAPHIC DATA FORM

DEMOGRAPHIC DATA FORM

Gender: M F

Age: ______

Major: ______________________

Class Standing: Freshman Sophomore Junior Senior Other

Handedness: Left-handed Right-handed

GPA:____________

SAT Verbal Score_________________

SAT Math Score___________________

How often do you work with personal computers?

_____ I've never worked with a personal computer
_____ Only a couple of times ever in my life
_____ Several times a year
_____ Several times a month
_____ Several times a week
_____ At least once a day, everyday
_____ For several hours everyday (over 4 hours a day)

Rate your experience with personal computers:
_____ Little or none
_____ Know a little; know Internet access, know some word processing and other software (e.g., Microsoft Word and Microsoft PowerPoint).
_____ Know quite a bit; know Internet access, know word processing well, used other software packages (e.g., Microsoft Access, FTP, WinZip), and have done some programming (e.g., HTML).
_____ Expert; know Internet access, word processing, other software, and have much experience with different programming languages (e.g., Flash, VB, C, and Java).

Do you currently or have you previously served in the military? YES NO

If yes, what is your current status? ACTIVE RESERVIST DISCHARGED

And what are/were your duties in the military?

____
____

Have you had any experience(s), which has made you familiar with military missions, equipment, and/or terminology (for example, are you involved in ROTC, have friends or relatives in the military/armed forces, etc.)? Please explain:

____

____

____

____

How many hours per week do you play video games? ______________

Please rate your experience with the following activities:

1. Playing virtual reality/first- person perspective video games (such as Doom, Quake, or Halo)

Not at all Experienced	Somewhat Experienced	Very Experienced
1	2	3

2. Playing third-person perspective or overview video games (such as Super Mario Brothers)

Not at all Experienced	Somewhat Experienced	Very Experienced
1	2	3

3. Doing sculpture, painting, drawing, or other visual arts

Not at all Experienced	Somewhat Experienced	Very Experienced
1	2	3

4. Constructing verbal arguments (such as debating or writing)

Not at all Experienced	Somewhat Experienced	Very Experienced
1	2	3

5. Solving word puzzles (such as crosswords)

Not at all Experienced	Somewhat Experienced	Very Experienced
1	2	3

6. Solving picture puzzles (such as hidden picture or jigsaw puzzles)

Not at all Experienced	Somewhat Experienced	Very Experienced
1	2	3

APPENDIX C : VISUALIZER-VERBALIZER QUESTIONNAIRES

Visual-Spatial Ability Rating (Mayer & Massa, 2003)

a. Please rate your verbal ability (check one):

______ Very High
______ Somewhat High
______ Average
______ Somewhat Low
______ Very Low

b. Please rate your spatial ability (check one):

______ Very High
______ Somewhat High
______ Average
______ Somewhat Low
______ Very Low

Santa Barbara Learning Style Questionnaire (SBLSQ; Mayer & Massa, 2003)

Please place a check mark in the corresponding box to indicate your level of agreement or disagreement.

a. I prefer to learn visually.

Strongly agree	Moderately agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Moderately disagree	Strongly disagree

b. I prefer to learn verbally.

Strongly agree	Moderately agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Moderately disagree	Strongly disagree

c. I am a visual learner.

Strongly agree	Moderately agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Moderately disagree	Strongly disagree

d. I am a verbal learner.

Strongly agree	Moderately agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Moderately disagree	Strongly disagree

e. I am good at learning from labeled pictures, illustrations, graphs, maps, and animations.

Strongly agree	Moderately agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Moderately disagree	Strongly disagree

f. I am good at learning from printed text.

Strongly agree	Moderately agree	Slightly agree	Neither agree nor disagree	Slightly disagree	Moderately disagree	Strongly disagree

Verbalizer-Visualizer Questionnaire (VVQ; Richardson, 1977)

Please indicate whether you believe the following statements are “True” or “False” by placing an X in the corresponding column.

		True	False
1	I enjoy doing work that requires the use of words.		
2	My daydreams are sometimes so vivid I feel as though I actually experience the scene.		
3	I enjoy learning new words.		
4	I can easily think of synonyms for words.		
5	My powers of imagination are higher than average.		
6	I seldom dream.		
7	I read rather slowly.		
8	I cannot generate a mental picture of a friend’s face when I close my eyes.		
9	I don’t believe that anyone can think in terms of mental pictures.		
10	I prefer to read instructions about how to do something rather than have someone show me.		
11	My dreams are extremely vivid.		
12	I have better than average fluency in using words.		
13	My daydreams are rather indistinct and hazy.		
14	I spend very little time attempting to increase my vocabulary.		
15	My thinking often consists of mental pictures or images.		

Participant ________

Multimedia Learning Preference Questionnaire**

Please read this text:

> Cool, moist air moves over a warmer surface and becomes heated. Warmed moist air near the earth's surface rises rapidly. As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud. The cloud's top extends above the freezing level. At this altitude, the air temperature is well below freezing so the upper portion of the cloud is composed of tiny ice crystals.

Suppose you need help on understanding the above text

You can click on one help icon and get this:

Help Screen 1

"water vapor" MEANS moisture in air that is in gas form such as in rising air before it condenses into a cloud

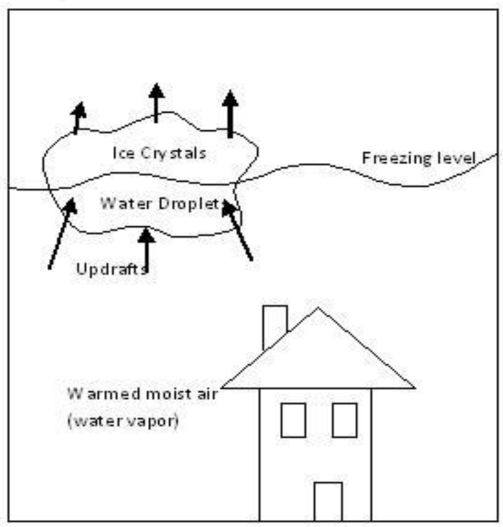
"water droplets" MEANS moisture in air that is in liquid form such as the part of a cloud below the freezing level

"ice crystals" MEANS moisture in air that is in solid form such as in the part of a cloud above the freezing level

"freezing level" MEANS at some point above the surface of the earth there is an imaginary line in the sky so that above the line water in a cloud will freeze into ice crystals and below the line water in a cloud will stay as water droplets

Or you can click on another help icon for this:

Help Screen 2



1. Warm moist air rises; water vapor condenses and forms a cloud.

Which of the two help screens do you prefer?

○	○	○	○	○	○	○
Strongly prefer 1	Moderately prefer 1	Slightly prefer 1	Equally like 1 and 2	Slightly prefer 2	Moderately prefer 2	Strongly prefer 2

Please read this text:

Eventually, the water droplets and ice crystals become too large to be suspended by the <u>updrafts</u>. As raindrops and ice crystals fall through the cloud, they drag some of the air in the cloud downward, producing <u>downdrafts</u>. When downdrafts strike the ground, they spread out in all directions, producing the gusts of cool wind people feel just before the start of rain.

Suppose you need help on understanding the above text

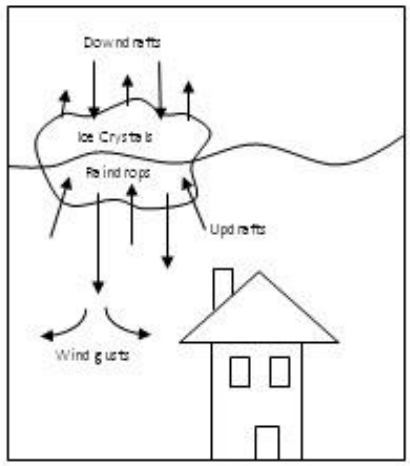
You can click on one help icon and get this:

Help Screen 1

"updraft" MEANS that a body of air is moving upward because it is warmer than surrounding air "downdraft" MEANS that a body of air is moving downward because it is cooler than the surrounding air

Or you can click on another help icon for this:

Help Screen 2



2. Raindrops and ice crystals drag air downward

Which of the two help screens do you prefer?

○	○	○	○	○	○	○
Strongly prefer 1	Moderately prefer 1	Slightly prefer 1	Equally like 1 and 2	Slightly prefer 2	Moderately prefer 2	Strongly prefer 2

Please read this text:

> Within the cloud, the rising and falling air currents cause electrical charges to build. The charge results from the collision of the cloud's rising water droplets against heavier falling pieces of ice. The negatively-charged particles fall to the bottom of the cloud and most of the positively-charged particles rise to the top.

Suppose you need help on understanding the above text

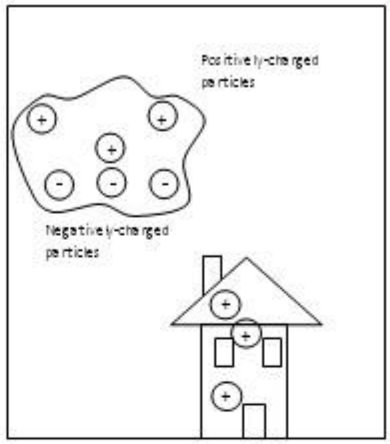
You can click on one help icon and get this:

Help Screen 1

> "electrical charge" MEANS the negatively-charged particles and positively-charged particles in material have been separated
>
> "negatively-charged particles" MEANS a part of the material in clouds that has a negative electrical charge, which normally is attracted to positively-charged particles
>
> "positively-charged particle" MEANS a part of the material in clouds that has a positive electrical charge, which normally is attracted to negatively-charged particles

Or you can click on another help icon for this:

Help Screen 2



3. Negatively charged particles fall to the bottom of the cloud.

Which of the two help screens do you prefer?

○	○	○	○	○	○	○
Strongly prefer 1	Moderately prefer 1	Slightly prefer 1	Equally like 1 and 2	Slightly prefer 2	Moderately prefer 2	Strongly prefer 2

Please read this text:

A stepped leader of negative charges moves downward in a series of zig-zag steps. It nears the ground. A positively charged leader travels upward from such objects as trees and buildings. The two leaders generally meet about 165 feet above the ground. Negatively-charged particles then rush from the cloud to the ground along the path created by the leaders. It is not very bright

Suppose you need help on understanding the above text

You can click on one help icon and get this:

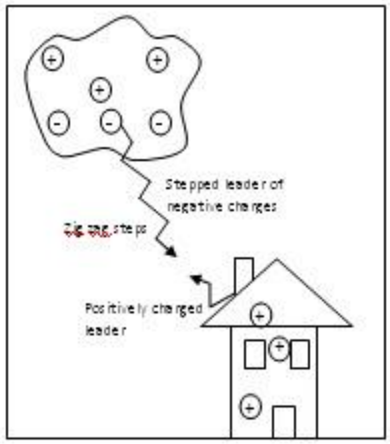
Help Screen 1

"stepped leader of negative charges" MEANS that negatively-charged particles from the bottom of the cloud move downward toward the positively-charged particles in objects on the earth's surface.

"positively charged leader" MEANS that positively-charged particles from objects on the earth's surface move upward toward the stepped leader of negative charges

Or you can click on another help icon for this:

Help Screen 2



4. Two leaders meet, negatively charged particles rush from the cloud to the ground

Which of the two help screens do you prefer?

○	○	○	○	○	○	○
Strongly prefer 1	Moderately prefer 1	Slightly prefer 1	Equally like 1 and 2	Slightly prefer 2	Moderately prefer 2	Strongly prefer 2

Please read this text:

As the leader stroke nears the ground, it induces an opposite charge, so positively-charged particles from the ground rush upward along the same path. This upward motion of current is the return stroke. It produces the bright light that people notice as a flash of lightning.

Suppose you need help on understanding the above text

You can click on one help icon and get this:

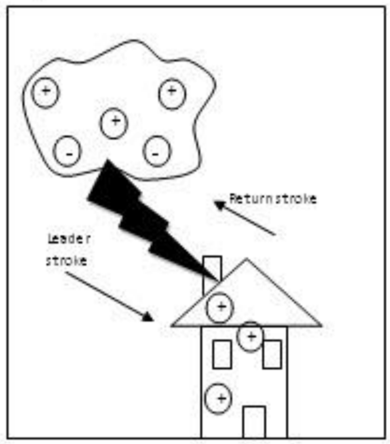
Help Screen 1

"leader stroke" MEANS that negatively-charged particles travel all the way from the cloud to the ground

"return stroke" MEANS that positively-charged particles travel all the way from the ground to the cloud

Or you can click on another help icon for this:

Help Screen 2



5. Positively charged particles from the ground rush upward along the same path.

Which of the two help screens do you prefer?

○	○	○	○	○	○	○
Strongly prefer 1	Moderately prefer 1	Slightly prefer 1	Equally like 1 and 2	Slightly prefer 2	Moderately prefer 2	Strongly prefer 2

APPENDIX D: KNOWLEDGE TEST

FOPC Sim Quiz

Please select the correct answer.

1. Which of the following is **not** one of the overarching rules of this simulation?

    a. Follow the prioritization rules.
    b. Correctly identify targets.
    c. Select effective ammunition types.
    d. Neutralize targets that move past your position.

2. Which of the following correctly describes how to change from tool to tool?
    a. Use the scroll wheel on the mouse or brackets on the keyboard
    b. Right click the mouse
    c. Left click the mouse

3. Which tool is used to determine the distance of a target?
    a. compass b. CFF sheet c. laser range finder

4. Which tool is used to determine the direction of a target?
    a. compass b. CFF sheet c. laser range finder

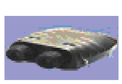
5. Which tool is selected to input the information for a CFF?
    a. compass b. CFF sheet c. laser range finder

6. After all information has been entered into the CFF sheet, what button do you press to send the transmission?
    a. Continue b. K c. Enter

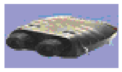
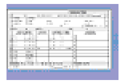
7. When you receive a Say Again, what does that indicate?
    a. incorrect/incomplete text entry
    b. select Continue
    c. k wasn't pressed

8. After the shots make impact, how do you clear the information in the CFF sheet?
    a. Mouse scroll bar
    b. Select Continue
    c. Hit escape

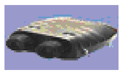
9. Which of the following pictures denotes the compass?

    a.  b.  c. 

10. Which of the following pictures denotes the laser range finder?

a.  b.  c. 

11. Which of the following pictures denotes the CFF sheet?

a.  b.  c. 

12. Which of the following correctly describes how to get a target's range using the laser range finder?

   a. Scroll to the laser range finder, right click to zoom it, left click to get distance, right click or escape to get out
   b. Scroll to the laser range finder and right click
   c. Scroll to the laser range finder, left click to zoom it, right click to get distance, left click or escape to get out

13. How will you know if a target has been neutralized?
   a. Black smoke
   b. It stops moving
   c. Both of the above

14. Should you fire on a target once it's been neutralized?
   a. yes b. no

15. How many meters per second does a tank travel?
   b. 200
   c. 25
   d. 10
   e. 8

16. How many seconds does it take for a round to land once the CFF has been completed?
   a. 200
   b. 25
   c. 10
   d. 8

17. When engaging a T-72 what type of ammunition is 100% effective?
   a. H E Quick b. ICM c. VT

APPENDIX E: WORKLOAD MEASURE

Multiple Resources Questionnaire

The purpose of this questionnaire is to characterize the nature of the mental processes used in the task with which you have become familiar. Below are the names and descriptions of several mental processes. Please read carefully so that you understand the nature of the process. Then rate the task on the extent to which it uses each process, using the following scale.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

Important:
All parts of a process definition should be satisfied for it to be judged as having been used. For example, recognizing geometric figures presented visually should not lead you to judge that the "tactile figural" process was used, just because figures were involved. For that process to be used, figures would need to be processed tactilely (i.e., using the sense of touch).
Please judge the task as a **whole**, averaged over the time you performed it. If a certain process was used at one point in the task and not another, your rating should **not** reflect "peak usage" but should instead reflect **average** usage over the entire length of the task.

1. **Auditory emotional process** – Required judgments of emotion (e.g., tone of voice or musical mood) presented through the sense of hearing.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

2. **Auditory linguistic process** – Required recognition of words, syllables, or other verbal parts of speech presented through the sense of hearing.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

3. **Facial figural process** – Required recognition of faces, or of the emotions shown on faces, presented through the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

4. **Facial motive process** – Required movement of own face muscles, unconnected to speech or the expression of emotion.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

5. **Manual process –** Required movement of the arms, hands, and/or fingers.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

6. **Short-term memory process –** Required remembering of information for a period of time ranging from a couple of seconds to half a minute.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

7. **Spatial attentive process –** Required focusing of attention on a location, using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

8. **Spatial categorical process –** Required judgment of simple left-versus-right or up-versus-down relationships, without consideration of precise location, using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

9. **Spatial concentrative process –** Required judgment of how tightly spaced are numerous visual object or forms.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

10. **Spatial emergent process –** Required "picking out" of a form or object from a highly cluttered or confusing background, using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

11. **Spatial positional process –** Required recognition of a precise location as differing from other locations, using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

12. **Spatial quantitative process –** Required judgment of a numerical quantity based on a nonverbal, nondigital representation (for example, bar graphs or small clusters of items), using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

13. **Tactile figural process –** Required recognition or judgment of shapes (figures), using the sense of touch.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

14. **Visual lexical process –** Required recognition of words, letters, or digits, using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

15. **Visual phonetic process –** Required detailed analysis of the sounds of words, letters, or digits, presented using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

16. **Visual temporal process –** Required judgment of time intervals, or of the timing of events, using the sense of vision.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

17. **Vocal process –** Required use of your voice.

No usage	Light usage	Moderate usage	Heavy usage	Extreme usage
0	1	2	3	4

APPENDIX F : FEEDBACK MANIPULATION CHECK QUESTIONNAIRE

1. On this scenario, how many times did you choose a lower priority target?

- **a. 0-2**
- **b. 3-4**
- **c. 5-7**
- **d. 8 or more**

2. What information did the feedback provide when you selected a lower priority the target?

____
____
____
____
____
____
____

3. On this scenario, how many times did your munitions miss the target?

- **a. 0-2**
- **b. 3-4**
- **c. 5-7**
- **d. 8 or more**

4. What information did the feedback provide when your munitions missed the target?

____
____
____
____
____

APPENDIX G : REACTIONS QUESTIONNAIRES

Feedback Reactions Questionnaire

Did you receive feedback during or after each scenario?
DURING AFTER I did not receive feedback
Was your feedback presented with text or spoken?
TEXT SPOKEN I did not receive feedback
If you received feedback, please continue. If not, skip to question 12.

Please think about the feedback you received during the first phase of training and indicate on the scale from 1-6 your level of agreement or disagreement with the following statements.		Strongly Agree					Strongly Disagree
		1	**2**	**3**	**4**	**5**	**6**
1.	The feedback I received was easy to understand.						
2.	I believe that the feedback I received correctly diagnosed the errors I was making.						
3.	I believe that the feedback I received helped me to improve my performance on the subsequent trial.						
4.	I believe that the feedback I received focused my attention on learning strategies to perform this task better.						
5.	I believe that the feedback I received focused my attention toward the performance level I should obtain.						
6.	I believe that the feedback I received could have been more useful.						
7.	It seemed like I received the same feedback over and over.						
8.	I believe that the feedback I received did not accurately reflect my performance.						
9.	I ignored and made no attempt to use the feedback I had received.						
10.	I believe that the feedback I received provided me with effective strategies to help me perform better.						
11.	I believe that the feedback I received helped me generate my own strategies to help me perform better.						
Skip to Question 16							

ONLY ANSWER THESE QUESTIONS IF YOU DID NOT RECEIVE FEEDBACK. Please indicate on the scale from 1-6 your level of agreement or disagreement with the following statements.	Strongly Agree					Strongly Disagree

		1	2	3	4	5	6
12.	I believe that feedback would have helped me improve my performance.						
13.	I would have liked to have received feedback on my performance.						
14.	I believe that having feedback would have motivated me more.						
15.	I believe that having feedback would have increased my confidence more.						

16. I have the following additional comments I would like to make concerning the feedback I was just provided with during this experiment.

____

____

____

APPENDIX H: DEBRIEF FORM

Debrief

Thank you for participating in today's experiment. You have participated in a study where participants play scenarios and receive different types of feedback at different times (during scenario or after scenario). Training is a crucial component in the military, particularly with the FO task, because serious incidents can occur from incorrect identifications of targets, incorrect munition choices, and incorrect prioritization. We are interested in automating the training process as much as possible in the future. A means to achieve this is to automatically analyze a trainee's performance data and provide feedback. This can be accomplished during training performance or delayed until after the completion of the scenario. We are interested in finding out which intervention strategy is best for providing feedback. We will use your data on the FO task to see which intervention look the most promising for the future of automatic feedback. We are evaluating the presentation and timing of feedback. We are not evaluating you.

If you are interested in more information about this project, we will be happy to provide you with an abbreviated abstract of the results once the data collection is finished. Let us know before you leave if you want to receive an abstract.

Thank you for your time!

APPENDIX I: FEEDBACK TEMPLATES

Outcome feedback statements for correct actions

Mission # X, Target #X
You are correct! You successfully chose the highest priority target.
You are correct! You were X meters from the target.

Outcome feedback statements for incorrect actions

Mission # X, Target #X
Incorrect. You performed the mission on a target that was not the highest priority target.
Incorrect. You were X meters from the target. Your shot did not disable the target.

Outcome feedback template for targets that are never neutralized by scenario completion

You failed to neutralize 3 of the 8 targets in the scenario. The following targets were not disabled: # 1, # 2, # 3

Outcome feedback template for targets that are not neutralized and come within 100 meters

Failed to neutralize target #X
Incorrect. You missed an opportunity to perform a mission on the highest priority target.
Incorrect. You did not disable the highest priority target.

Process feedback statements for correct actions

Mission # X, Target #X
Continue locating and comparing the priority of targets before performing a mission.
For stationary targets--Continue correctly using the compass and laser range finder to determine a target's distance.
OR
For moving targets--Continue correctly using the laser range finder to project the target's location into the future.

Process feedback statements for incorrect actions

Mission # X, Target #X
Be sure to locate and compare the priority of targets before performing a mission.
For stationary targets -- Be sure to right click the mouse when using the laser range finder to determine distance.
For moving targets--Your munition landed behind the target. It will be easier to hit a moving target if you enter all other information in the CFF sheet BEFORE checking and entering the range.

For moving targets --Your munition landed in front of the target. It will be easier to hit a moving target if you enter all other information in the CFF sheet BEFORE checking and entering the range.

Process feedback template for targets that are never neutralized by scenario completion

Failed to disable the following targets: # 1, # 2, # 3
Be sure to locate and compare the priority of targets before performing a mission.
Use the compass and the laser range finder to determine a target's location.

Process feedback template for targets that are not neutralized and come within 100 meters

Failed to neutralize target #X
Be sure to locate and compare the priority of targets before performing a mission.
It will be easier to hit a moving target if you enter all other information in the CFF sheet BEFORE checking and entering the range.

APPENDIX J: IRB APPROVAL DOCUMENTS

DEPARTMENT OF THE NAVY
NAVAL AIR WARFARE CENTER TRAINING SYSTEMS DIVISION
12350 RESEARCH PARKWAY
ORLANDO, FLORIDA 32826-3275

IN REPLY REFER TO
3900
Ser HRPP/185
JUN 04 2010

From: Commanding Officer, Naval Air Warfare Center Training Systems Division
To: Wendi Van Buskirk

Subj: APPROVAL CERTIFICATION FOR PROTOCOL, TSD 185, INVESTIGATING THE RELATIONSHIP BETWEEN FEEDBACK TIMING, CONTENT AND MODALITY IN SCENARIO-BASED TRAINING EXERCISES

Ref: (a) 32 Code of Federal Regulation 219

Encl: (1) Protocol Information Document for TSD 185

1. In accordance with reference (a), a review of Protocol TSD 185, Investigating the Relationship between Feedback Timing, Content and Modality in Scenario-based Training Exercises, was conducted by the Naval Air Warfare Center Training Systems Division (NAWCTSD) Committee for the Protection of Human Subjects (CPHS) Chair on 25 May 2010. It was determined that the risk of the study is minimal and that the study is exempt. The CPHS Chair recommends approval of the protocol as written.

2. Protocol TSD 185 is hereby approved. Your protocol expires on 24 May 2011, one year from the date of the Exempt Review. Should there be a need to extend this protocol beyond the one-year approval period, you are responsible for submitting a Continuing Review (CR) Status Report and supporting documentation to the NAWCTSD CPHS in sufficient time to allow for appropriate review and re-approval before the end of the current approval period. Human subject research shall not be conducted outside of an approval period. If the approval period expires, Federal law requires that the CPHS temporarily suspend the study and that work involving human subjects temporarily cease. CR must be properly completed and re-approval granted before the end of the current approval period in order to avoid interruption of the research. The CPHS will send a reminder to you approximately two months prior to the expiration of your protocol.

3. Please provide a copy of enclosure (1) to each voluntary subject prior to their participation in the study.

Subj: APPROVAL CERTIFICATION FOR PROTOCOL, TSD 185, INVESTIGATING THE RELATIONSHIP BETWEEN FEEDBACK TIME, CONTENT, AND MODALITY IN SCENARIO-BASED TRAINING EXERCISES

4. A Final Report is due to the NAWCTSD CPHS no more than two months after data analysis is complete. The CPHS considers a research study to be active until data analysis is complete and subject to continuing review.

5. Should you have any questions, please contact the NAWCTSD CPHS Administrator at (407) 380-4320.

H. M. ROBINSON

2

APPROVED JUN 0 4 2010. This document may NOT be used after MAY 2 4 2011 .

APPENDIX A.
PROTOCOL INFORMATION DOCUMENT
(EXEMPT STUDY)

1. You are being asked to voluntarily participate in a research study entitled Investigating the Relationship between Feedback Timing, Content, and Modality in Scenario based Training Exercises. You will be asked to learn a military call for fire task within a simulator test-bed. Specifically, you will be asked to participate in a series of training sessions that will teach you to effectively perform the call for fire task within the simulator test-bed. You will be thoroughly familiarized with the simulator test-bed. Throughout the experimental session, you will also be asked to fill out several questionaires and forms concerning your performance within the simulator. Your performance on the simulated call for fire task will be objectively and internally recorded by the simulator itself. You must be 18 years old to participate in this study. The researchers expect that there will be approximately 99 research subjects participating in this study. It is expected that this study will take 3 hours. Breaks will be scheduled during this study; however, you may take breaks as needed.

2. The investigators believe that the risks or discomforts to you are as follows: There are no known risks. This study involves the use of typical video and/or computer games. If you have a history of seizures YOU WILL BE UNABLE TO PARTICIPATE in this study. Additionally, there are no known or expected risks to a pregnant woman, the embryo or fetus.

3. The benefits that you may expect from your participation in this study are that you will receive $30 for 3 hours of participation.

4. Your confidentiality during the study will be ensured by assigning you a coded identification number. Your name will not be directly associated with any data. Any subject identification keys will be destroyed at the end of the study. This procedure will insure that your personal data cannot be used in any way that might impact your career, academic progress, or standing in your respective professional or educational communities. Please understand that all personal data or information (such as demographic data/performance data) will be secured under lock and key until destroyed in accordance with SECNAV M-5210.1 and as required by 32 CFR 219.111(a)(7).

5. Should you have questions concerning the research described in this Protocol Information Document or questions concerning research-related injury, please contact the Principal Investigator listed below. Additionally, if you so desire, you may contact the Principal Investigator for a copy of any publication resulting from this study. If you have questions concerning your rights as a research subject, please contact the CPHS Chair or Vice Chair (see contact info provided below).

a. Principal Investigator: Wendi Van Buskirk
Activity: NAVAIR Orlando
Mailing Address: 12350 Research Parkway, Orlando, FL 32826
Code: 4651 Phone: 407-380-4558 E-mail: wendi.vanbuskirk@navy.mil

APPROVED JUN 04 2010. **This document may NOT be used after** MAY 24 2011.

b. Project Manager: Dr. Chris Hicks
Activity: Naval Air Systems Command
Mailing Address: 48110 Shaw Road, Bldg. 2187, Room 2172, Patuxent River, MD 20670
Code: 4.5 Phone: 301-342-9094 E-mail: chris.hicks@navy.mil

c. CPHS Chairman: Dr. Katrina Ricci
Activity: Naval Air Warfare Center Training Systems Division
Mailing Address: 12350 Research Parkway, Orlando, FL 32826-3275
Code: 4.6.5.1 Phone: (407) 380-4661 E-mail: Katrina.Ricci@navy.mil

6. Participation in this research study is voluntary. You may choose not to participate. If you decide to withdraw from further participation in this study, there will be no penalties. To ensure your safe and orderly withdrawal from the study, please inform the Principal Investigator listed in paragraph 5.a. of this document.

7. Your participation in this study may be stopped by the investigator at any time without your consent if it is believed the decision is in your best interest. There will be no penalty or loss of benefits to which you are otherwise entitled at the time your participation is stopped.

8. No out of pocket costs to you may result from your voluntary participation in this study.

9. Official government agencies may have a need to inspect the research records from this study, including your records, in order to fulfill their responsibilities.

10. A Privacy Act Statement concerning this research protocol is attached as Appendix B and will be provided to you prior to your participation in this study.

11. It is important that you understand what has been explained in this Protocol Information Document about your participation in this study. If you have any questions or concerns about this study and its related procedures and any risks that may be associated with your participation, please talk with the Principal Investigator listed in paragraph 5.a. All of your questions should be answered to your satisfaction prior to your participation.

LIST OF REFERENCES

Adepoju A.A & Elliott R.T. (1997). Comparison of different feedback procedures in second language vocabulary learning. *Journal of Behavioral Education, 7*(4), 477-495

Akamatsu, M., MacKenzie, I. S., Hasbroucq, T. (1995). A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, *38*(4), 816-827.

American Psychological Association (2002). *Ethical principles of psychologists and code of conduct*. Retrieved from http://www.apa.org/ethics/code2002.pdf

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 167-207.

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1971). Feedback procedures in programmed instruction. *Journal of Educational Psychology, 62*(2), 148-156.

Anderson, R. C., Kulhavy, R. W., & Andre, T. (1972). Conditions under which feedback facilitates learning from programmed lessons. *Journal of Educational Psychology, 63*(2), 186-188.

Astwood, R. S., Van Buskirk, W. L., Cornejo, J., & Dalton, J. (2008). The Impact of Different Feedback Types on Decision-Making in Simulation Based Training Environments. *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society* [CD-ROM].

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 11-127.

Baddeley. A. D. (2000). The episodic buffer: A new component of working memory? Trends *in Cognitive Sciences, 4*(11), 417-422.

Baddeley. A. D. (2001). Is working memory still working. *American Psychologist, 56*, 849-864.

Baddeley. A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake and P. Shah (Eds.) *Models of working memory: Mechanisms of active maintenance and executive control.* United Kingdom: Cambridge University Press.

Balzer, W. K., Hammer, L. B., Sumner, K. E., Birchenough, T. R., Martens, S. P., & Raymark, P. H. (1994). Effects of cognitive feedback components, display format, and elaboration on performance. *Organizational Behavior and Human Decision Processes, 58,* 369-385.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Bisantz, A. M. & Sharit, J. (1993). The effects of feedback on performance and retention of skill for a natural language interface. *Behavior and Information Technology*, *12*(1), 32-47.

Boles, D. B., Bursk, J. H., Phillips, J. B., Perdelwitz, J. R. (2007). Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). *Human Factors, 49*(1), 32-45

Bolton, A. E. (2006). *Immediate versus delayed feedback in simulation based training: Matching feedback delivery timing to the cognitive demands of the training exercise*. Unpublished doctoral dissertation, University of Central Florida, Orlando.

Bowers, C.A. & LaBarba, R.C. (1988). Sex differences in the lateralization of spatial abilities: A spatial component analysis of extreme group scores. *Brain and Cognition, 8*, 165-177.

Brackbill, Y., Bravos, A., & Starr, R. H. (1962). Delay-improved retention of a difficult task. *Journal of Comparative and Physiological Psychology, 55*, 947-952.

Buff, W. L. & Campbell, G. E. (2002). What to do or what not to do?: Identifying the content of effective feedback. *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*, (pp.2074-2078). Santa Monica, CA: HFES.

Clariana, R. B., Ross, S. M., & Morrison, G. R. (1991). The effects on learning of different feedback strategies using computer-assisted multiple-choice questions as instruction. *Educational Technology, Research, & Development, 39*, 5-17.

Clariana, R. B., Wagner, D., and Murphy, R. L. C. (2000). A connectionist description of feedback timing. *Educational Technology Research and Development*, 48, 5-11.

Cohen, J. & Cohen, P. (1983). *Applied Mulitple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: LEA.

Corbett, A. T. & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. *Proceedings of ACM CHI Conference on Human Factors in Computing Systems*, 245-252.

Corbett, A. T., Koedinger, K. R. & Anderson, J. R. (1997). Intelligent tutoring systems. In M.G. Helander, T.K. Landauer & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 849-74). Amsterdam: Elsevier.

Del Rey P, Shewokis PA. (1993). Appropriate KR summary lengths for learning timing tasks under conditions of high and low contextual interference. *Acta Psychologica*, 83, 1-12.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *Psychological Record, 53*, 533-548.

Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal, 33*, 87-105.

Farmer, E., van Rooji, J., Riemersma, J., Jorna, P., & Morall, J. (1999). *Handbook of Simulator-Based Training*. Burlington, VT: Ashgate.

Ferris, T., Hameed, S., Penfold, R. & Rao, N. (2007). Tactons in multitask environments: The interaction of presentation modality and processing code. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, 1481-1485.

Gaynor, P. (1981). The effect of feedback delay on retention of computer-based mathematical material. *Journal of Computer-Based Instruction, 8*(2), 28-34.

Geary, D.C. & DeSoto, M.C. (2001). Sex differences in spatial abilities among adults from the United States and China: Implications for an evolutionary theory. *Evolution and Cognition*, *7*(2), 172-177.

Gilman, D. A. (1970). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology, 60*, 503-508.

Guay, M., Salmoni, A., McIlwain, J. (1992). Summary knowledge of results for skill acquisition: Beyond Lavery and Schmidt. *Human Movement Science, 11*(6), 653-673.

Guthrie, E. R. (1935). *The Psychology of Learning*. New York: Harper.

Ilgen, D. R., Fisher, C. D., & Taylor, M. T. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*, 349-371.

Jonassen, D. H. & Grabowski, B. L. (1993). *Handbook of Individual Differences, Learning and Instruction.* Hillsdale, NJ: Lawrence Erlbaum.

Kirlik, A., Fisk, A. D., Walker, N., Rothrock, L. (1998). Feedback augmentation and part-task practice in training dynamic decision-making skills. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decision under stress: Implications for individual and team training* (pp 91-113). APA: Washington, DC.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions of performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254284.

Kozlowski, S. W. J., Toney, R. J., Mullins, M. E., Weissbein, D. A., Brown, K. G., & Bell, B. S. (2001). Developing adaptability: A theory for the design of integrated-embedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol.1, pp. 59-123). Amsterdam: JAI/Elsevier Science.

Kulhavy (1977). Feedback in written instruction. *Review of Educational Research, 47*, 211-232.

Kulhavy, R. W. & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology, 63*, 505-512.

Kulhavy, R. W., and Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279 - 308.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.

Landsberg, C., Van Buskirk, W**.,** Astwood, R. (2010). Does feedback type matter? Investigating the effectiveness of feedback content on performance outcomes. *Proceedingss of the 54th Annual Meeting of the Human Factors and Ergonomics Society* [CD-ROM].

Lesgold, A., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. In J. Regian and V. Shute (Eds.) *Cognitive Approaches to Automated Instruction*. Hillsdale, NJ: LEA.

Leutner, D. & Plass, J. L. (1998). Measuring learning styles with observational versus direct observation of preferential choice behavior in authentic learning situations: The visualizer/verbalizer behavior observation scale (VV-BOS). *Computers in Human Behavior, 14*(4), 543-557.

Locke, E. A., & Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice-Hall.

Mason, B., & Bruning, R. (2003). *Providing feedback in computer-based instruction: What the research tells us*. Retrieved from http://dwb.unl.edu/Edit/MB/MasonBruning.html

Mayer, R. E. (1999). Research-based principles for the design of instructional messages: The case of multimedia explanations. *Document Design, 1*, 7-20.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2002). *The Promise of Educational Psychology* (Vols. 1 & 2). Upper Saddle River, NJ: Prentice-Hall.

Mayer, R. E. & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95*(4), 833-841.

McLaughlin, A. C., Rogers, W. A., & Fisk, A. D. (2006). How effective feedback for training depends on learner resources and task demands. *Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society* [CD-ROM].

Moreno, R. & Mayer, R. E. (2000). A learner-centered approach to multimedia explanations: Deriving instructional design principles from cognitive theory. *Interactive Multimedia Electronic Journal of Computer-enhanced Learning, 2*(2). Retrieved from http://imej.wfu.edu/articles/2000/2/05/index.asp

Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Morrison, G. R., Ross, S. M., Gopalakrishnan, M., & Casey, J. (1995). The effects of feedback and incentives on achievement in computer-based instruction. *Contemporary Educational Psychology, 20*(1), 32-50.

Mory, E. (1992). The use of informational feedback in instruction: Implications for future research. *Educational Training Research and Development*, *40(3),* 5-20.

Neth, H., Khemlani, S. S., & Gray, W. D. (2008). Feedback design for the control of a dynamic multitasking system: Dissociating outcome feedback from control feedback. *Human Factors, 50*(4), 643-651.

Oser, R. L., Cannon-Bowers, J. A., Salas, E., &, Dwyer, D. J. (1999). Enhancing human performance in technology-rich environments: Guidelines for scenario-based training. *Human/technology interaction in complex systems, Vol. 9*, pp. 175-202. Stamford, CT: JAI Press, Inc.

Oser, R. L., Cannon-Bowers, J. A., Dwyer, D. J., & Salas, E. (1997). Establishing a learning environment for JSMIS: Challenges and considerations [CD-ROM]. *Proceedings of the 19th Annual Interservice/Industry Training, Simulation and Education Conference* (pp. 144-153). Orlando, FL: National Training Systems Association.

Panasuk, R. & LeBaron, J. (1999). Student Feedback: A Tool for Improving Instruction. *Education*, *120* (2), 356-368.

Parush, A. (2005). Speech-based interaction in multitask conditions: Impact of prompt modality. *Human Factors, 47*(3), 591-597.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles: Concept and evidence. *Psychological Science in the Public Interest*, *9*(3), 105-119.

Rhodenizer Van Duyne, L. R. (2001). *Sequencing Process and Outcome Feedback: Investigating the Role of Goal Orientation During Skill Acquisition*. Unpublished doctoral dissertation, University of Central Florida, Orlando.

Richardson, A. (1977). Verbalizer-Visualizer: A cognitive style dimension. *Journal of Mental Imagery*, 1, 109-126.

Roper, W. J. (1977). Feedback in computer assisted instruction. *Programmed Learning and Educational Technology, 14*(1), 43 - 49.

Rose, F. D., Attree, E. A., Brooks, B. M., Parslow, D. M., Penn, P. R., & Ambihaipahan, N. (2000). Training in virtual environments: Transfer to real world tasks and equivalence to real task training. *Ergonomics*, *43*(4), 494-511.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It's not how much you have but how you use it: Towards a rational use of simulation in aviation training. *International Journal of Aviation Psychology, 8,* 197-208.

Schmidt, R.A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In J. Requin & G.E. Stelmach (Eds.), *Tutorials in Motor Neuroscience* (pp. 59-75). Dordrecht: Kluwer Academic Publishers.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Schmidt, R.A., & Wulf, G. (1997). Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human Factors*, *39*, 509-525.

Schooler, L. J. & Anderson, J. R. (1990). The disruptive potential of immediate feedback. In

Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, 702-708, Cambridge, MA.

Shneiderman, B. (1988). *Designing the user interface: Strategies for effective human-computer interaction* (3rd ed.). Addison Wesley Longman: Reading, MA.

Shute, V. (2007). *Focus on formative feedback* (RR-07-11). Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-07-11.pdf

Smither, J. W., Wohlers, A. J., & London, M. (1995). A field study of reactions to normative versus individualized upward feedback. *Group & Organization Management, 20,* 61-89.

Sturges, P. T. (1969). Verbal retention as a function of the informativeness and delay of informative feedback. *Journal of Educational Psychology, 60*, 11-14.

Sturges, P. T. (1972). Information delay and retention: Effect of information in feedback and tests. *Journal of Educational Psychology, 63*, 32-43.

Sturges, P. T. (1978). Delay of informative feedback in computer-assisted testing. *Journal of Educational Psychology, 70*, 378-387.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Thorndike, E. (1913). *Educational Psychology: The Psychology of Learning*. New York: Teachers College Press.

Webb, J. M., Stock, W. A., & McCarthy, M. T. (1994). The effects of feedback timing on learning facts: The role of response confidence. *Contemporary Educational Psychology, 19,* 251-265.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention* (pp. 63–102). San Diego, CA: Academic Press.

Wickens, C. D. (2008). Multiple resources and mental workload [Special issue]. *Human Factors, 50*(3), 447-455.

Wickens, C. D., & Hollands, J. (2000). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Zolna, J.S., Catambrone, R. (2007). The effects of study time and presentation modality on learning. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting*, 512- 515.